statistics. Interestingly, the ER model can be characterized as a special case of MRG, where all the parameters except for $\eta$ are set to zero.

MRGs are known to satisfy Definition 2 [Robins et al., 2007]. Therefore, under the probability model defined by Eq. 10, we have that, for any pair of nodes $\{u, v\}$ and any subgraph sample $G_S$ from $G$, $P(X_{uv} \mid \mathcal{A}(G_S) \setminus \{X_{uv}\}; \theta) = P(X_{uv} \mid N_S(X_{uv}); \theta)$, where $N_S(X_{uv})$ is the set of all variables $X_{wz}$ from $G_S$ such that $|\{u, z\} \cap \{u, v\}| = 1$. Note that while Definition 2 specifies a general class of random graph models, MRGs in the strict sense refer to the class of ERGs defined above.

### 4.1.2 Higher-Order Models

Higher-order ERG models (HRGs) are readily obtained from MRGs by adding $k$-triangle counts (for $k \geq 1$) to the Boltzmann distribution of Eq. 10 [Robins et al., 2007]. Moreover, in order to avoid fixing in advance the maximum value of the $k$ parameter, a general formulation has been proposed for HRGs through the alternating $k$-star and the alternating $k$-triangle statistics [Snijders et al., 2006], obtaining the following probability model:

$$P(G; \theta) = \frac{1}{Z} \exp \left\{ \eta E(G) + \sigma S^*(G) + \tau T^*(G) \right\}$$ (11)

where, if $n$ is the number of nodes in $G$ and $S_k$ and $T_k$ are the usual $k$-stars and $k$-triangle counts, $S^*(G)$ and $T^*(G)$ are defined as $S^*(G) = \sum_{k=1}^{n-1} (-1)^k \frac{S_k}{k!}$ and $T^*(G) = \sum_{k=2}^{n-2} (-1)^k \frac{T_k}{k!}$ (with $\rho \geq 1$ acting as a sort of regularization parameter).

It is known that, under the distribution given by Eq. 11, $P(X_{uv} \mid \mathcal{A}(G_S) \setminus \{X_{uv}\}; \theta) = P(X_{uv} \mid N_S(X_{uv}); \theta)$ [Robins et al., 2007], where $N_S(X_{uv})$ is the set containing any $X_{wz}$ from $G_S$ such that $X_{wz} \neq X_{uv}$, and, for at least one edge $\{s, t\}$, we have that $|\{s, t\} \cap \{u, v\}| = 1$ and $|\{s, t\} \cap \{w, z\}| \neq 0$. Clearly, $N_S(X_{uv}) \subseteq N_S(X_{uv})$, which is why HRGs are ‘higher-order’ than MRGs.

### 4.2 Watts-Strogatz Model

The WS network model [Watts and Strogatz, 1998] defines a random network as a regular ring lattice which is randomly ‘rewired’ so as to introduce a certain amount of disorder, which typically leads to small-world phenomena. Given nodes $u_1, \ldots, u_n$, a WS network is generated by constructing a regular ring lattice such that each node is connected to exactly $2\delta$ other nodes. Network edges are then scanned sequentially, and each one of them is rewired with probability $\beta$, where, if $i < j$, rewiring an edge $\{u_i, u_j\}$ means to replace it with another edge $\{u_i, u_k\}$ such that $k \neq i$ and $u_k$ is chosen uniformly at random from the set of all nodes that are not already linked to $u_i$.

Interestingly, the degree distribution corresponding to a WS network $G$ with parameters $\delta$ and $\beta$ takes the following form [Barrat and Weigt, 2000], for any degree $k \geq \delta$:

$$P(k) = \sum_{i=0}^{I} \binom{\delta}{i} (1 - \beta)^i \beta^{k-i} \left(\frac{\delta - i}{k - i}\right)! \exp(-\beta\delta)$$ (12)

where $I = \min\{k - \delta, \delta\}$. Given Eq. 12, we model the conditional distribution of a variable $X_{uv}$ given the remainder of $G$ through the following quantity:

$$P(X_{uv} \mid \mathcal{A}(G) \setminus \{X_{uv}\}) = \frac{P(d_u(G)) P(d_v(G))}{\sum_{x_{uv}} P(d_u(G_{x_{uv}})) P(d_v(G_{x_{uv}}))}$$ (13)

where $d_u(G)$ denotes the degree of node $u$ in $G$, and each $P$ is implicitly conditional on the values of $\delta$ and $\beta$. We refer to the conditional random graph model specified in Eq. 13 as the conditional Watts-Strogatz (CWS) model.

For the CWS model, the following proposition follows straightforwardly from Eq. 13:

**Proposition 4.** The dependence structure of CWS models is Markovian.

A simple estimate of the $\delta$ parameter for a data sample $\mathcal{D} = \{(x_{u_1v_1}, G_{u_1v_1}), \ldots, (x_{u_nv_n}, G_{u_nv_n})\}$ is the following:

$$\hat{\delta} = \frac{1}{2n} \sum_{i=1}^{n} d_u(G_{u_i}) + d_v(G_{u_i})$$ (14)

On the other hand, a simple strategy for estimating the rewiring probability by a maximum likelihood approach consists in parameterizing it as a sigmoid function $\beta = \frac{1}{1 + \exp(-\theta_{\beta})}$, where $\theta_{\beta}$ can be optimized by exploiting the derivative $\frac{\partial}{\partial \beta} \sum_{i=1}^{n} \log P(x_{u_i}, \mathcal{A}(G_{u_i}) \setminus \{X_{uv}\}; \delta, \theta_{\beta})$.

### 4.3 Barabási-Albert Model

The BA model was originally proposed for explaining the scale-free degree distributions often observed in real-world networks [Barabási and Albert, 1999]. In the BA model, the probability $P(u)$ of linking to any particular node $u$ in a network $G = (V, E)$ takes the form $P(u) = \frac{d_u(G)}{\sum_{v \in V} d_v(G)}$ [Albert and Barabási, 2002], where $\alpha$ is a real-valued parameter affecting the shape of the degree distribution. Given $P(u)$, we can use the following expression to characterize the conditional probability of observing edge $\{u, v\}$ given the