Controlled Markovian dynamics of graphs: unbiased generation of random graphs with prescribed topological properties

ES Roberts
Randall Division of Cell and Molecular Biophysics
King's College London
Email: ekaterina.roberts@kcl.ac.uk

A Annibale
Dept. of Mathematics
King's College London
alessia.annibale@kcl.ac.uk

ACC Coolen
Dept. of Mathematics and Randall Division of Cell and Molecular Biophysics
King's College London
ton.coolen@kcl.ac.uk

***Abstract*—We analyze the properties of degree-preserving Markov chains based on elementary edge switchings in undirected and directed graphs. We give exact yet simple formulas for the mobility of a graph (the number of possible moves) in terms of its adjacency matrix. This formula allows us to define acceptance probabilities for edge switchings, such that the Markov chains become controlled Glauber-type detailed balance processes, designed to evolve to any required invariant measure (representing the asymptotic frequencies with which the allowed graphs are visited during the process).**

I. Introduction

Sampling uniformly the space of graphs with prescribed macroscopic properties has become a prominent problem in many application areas where tailored graph ensembles are used as proxies or null models for real complex networks. Quantities measured in real networks are often compared with the values these quantities take in their randomised counterparts. Ensembles of randomised networks allow one to put error bars on these values, and therefore to identify which topological features of real networks deviate significantly from the null model. Such features are likely to reflect e.g. design principles or evolutionary history of the network. Ensembles of randomised graphs are to be generated numerically, and each graph realisation should be produced with a probability proportional to a prescribed statistical weight (often taken to be uniform) of the graph, for the analysis to be unbiased. At the level where the constraints of the randomized graphs involve only the simplest quantities, i.e. the degree sequence, even uniform generation of such graphs is known to be a non-trivial problem [1]–[4]. One classical algorithm for generating random networks with prescribed degrees [5] assigns to each node a number of 'edge stubs' equal to its desired degree, and joins iteratively pairs of randomly picked stubs to form a link. A drawback of this algorithm is the need for rejection of forbidden graphs (those with multiple edges or self-loops), which can lead to biased sampling [6]. A second popular method for generating random graphs with a given degree sequence is 'edge swapping', which involves successions of ergodic graph randomising moves that leave all degrees invariant [7], [8]. However, naive accept-all edge swapping will again cause sampling biases. The reason is that the *number* of edge swaps that can be executed is not a constant, it depends on the graph $\boldsymbol{c}$ at hand; graphs which allow for many moves will be generated more often. Any bias in the sampling of graphs invalidates their use as null models, so one is forced to mistrust all papers in which observations in real graphs have been tested against null models generated either via the 'stubs' method or via randomisation by 'accept-all edge swapping'. This situation can only be remedied via a systematic study of stochastic Markovian graph dynamics, which is the topic of this paper. We determine the appropriate adjustment to the probability of accepting a randomly chosen proposed edge swap for the process to visit each graph configuration with the same probability. This can be done for nondirected and directed graphs. We will also show that our method can be used to generate graphs with prescribed degree correlations.

II. Generating random graphs via Markov chains

A general and exact method for generating graphs from the set $G[\boldsymbol{k}] = \{\boldsymbol{c} \in G|\ \boldsymbol{k}(\boldsymbol{c}) = \boldsymbol{k}\}$ randomly (where $\boldsymbol{k}$ denotes the degree sequence, or the joint in- and out-degree sequence of the graph), with specified probabilities $p(\boldsymbol{c}) = Z^{-1}\exp[-H(\boldsymbol{c})]$ was developed in [9]. It has the form of a Markov chain:

$$\forall \boldsymbol{c} \in G[\boldsymbol{k}]: \quad p_{t+1}(\boldsymbol{c}) = \sum_{\boldsymbol{c}' \in G[\boldsymbol{k}]} W(\boldsymbol{c}|\boldsymbol{c}')p_t(\boldsymbol{c}') \tag{1}$$

Here $p_t(\boldsymbol{c})$ is the probability of observing graph $\boldsymbol{c}$ at time t in the process, and $W(\boldsymbol{c}|\boldsymbol{c}')$ is the one-step transition probability from graph $\boldsymbol{c}'$ to $\boldsymbol{c}$. For any set Φ of *ergodic* reversible elementary moves $F : G[\boldsymbol{k}] \to G[\boldsymbol{k}]$ we can choose transition probabilities of the form

$$W(\boldsymbol{c}|\boldsymbol{c}') = \sum_{F\in\Phi} q(F|\boldsymbol{c}')\Big[\delta_{\boldsymbol{c},F\boldsymbol{c}'}A(F\boldsymbol{c}'|\boldsymbol{c}') + \delta_{\boldsymbol{c},\boldsymbol{c}'}[1-A(F\boldsymbol{c}'|\boldsymbol{c}')]\Big] \tag{2}$$

The interpretation is as follows. At each step a candidate move $F \in \Phi$ is drawn with probability $q(F|\boldsymbol{c}')$, where $\boldsymbol{c}'$ denotes the current graph. This move is accepted (and the move $\boldsymbol{c}' \to$

$\boldsymbol{c} = F\boldsymbol{c}'$ executed) with probability $A(F\boldsymbol{c}'|\boldsymbol{c}') \in [0,1]$, which depends on the current graph $\boldsymbol{c}'$ and the proposed new graph $F\boldsymbol{c}'$. If the move is rejected, which happens with probability $1-A(F\boldsymbol{c}'|\boldsymbol{c}')$, the system stays in $\boldsymbol{c}'$. We may always exclude from Φ the identity operation. One can prove that the process (1) will converge towards the equilibrium measure $p_\infty(\boldsymbol{c}) = Z^{-1}\exp[-H(\boldsymbol{c})]$ upon making in (2) the choices

$$q(F|\boldsymbol{c}) = I_F(\boldsymbol{c})/n(\boldsymbol{c}) \tag{3}$$

$$A(\boldsymbol{c}|\boldsymbol{c}') = \frac{n(\boldsymbol{c}')\mathrm{e}^{-\frac{1}{2}[H(\boldsymbol{C})-H(\boldsymbol{C}')]}}{n(\boldsymbol{c}')e^{-\frac{1}{2}[H(\boldsymbol{C})-H(\boldsymbol{C}')]}+n(\boldsymbol{c})\mathrm{e}^{\frac{1}{2}[H(\boldsymbol{C})-H(\boldsymbol{C}')]}} \tag{4}$$

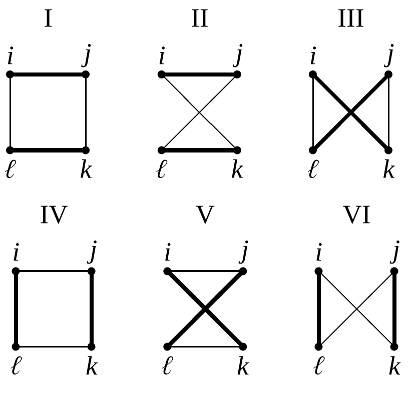
Here $I_F(\boldsymbol{c}) = 1$ if the move F can act on graph $\boldsymbol{c}$, $I_F(\boldsymbol{c}) = 0$ otherwise, and $n(\boldsymbol{c})$ denotes the total number of moves that can act on a graph $\boldsymbol{c}$ (the 'mobility' of state $\boldsymbol{c}$):

$$n(\boldsymbol{c}) = \sum_{F\in\Phi} I_F(\boldsymbol{c}). \tag{5}$$

III. Degree-constrained dynamics of nondirected graphs

We first apply our results to algorithms that randomise undirected graphs, while conserving all degrees, by application of edge swaps that act on quadruplets of nodes and their mutual links. Such moves were shown to be ergodic, i.e. any two graphs with the same degree sequence can be connected by a finite number ofsuccessive edge swaps [7], [8].

Let us define the set $Q = \{(i,j,k,\ell) \in \{1,\ldots,N\}^4 |\ i<j<k<\ell\}$ of all ordered node quadruplets. The possible edge swaps to act on (i,j,k,ℓ) are the following, with thick lines indicating existing links and thin lines indicating absent links that will be swapped with the existing ones, and where (IV, V, VI) are the inverses of (I, II, III):



We group the edge swaps into three pairs (I,IV), (II,V), and (III,VI), and label all three resulting auto-invertible operations for each ordered quadruple (i,j,k,ℓ) with a subscript α. Our auto-invertible edge swaps are now written as $F_{ijk\ell;\alpha}$, with $i<j<k<\ell$ and $\alpha \in \{1,2,3\}$. We define associated indicator functions $I_{ijk\ell;\alpha}(\boldsymbol{c}) \in \{0,1\}$ that detect whether (1) or not (0) the edge swap $F_{ijk\ell;\alpha}$ can act on state $\boldsymbol{c}$, so

$$I_{ijk\ell;1}(\boldsymbol{c}) = c_{ij}c_{k\ell}(1-c_{i\ell})(1-c_{kj}) + (1-c_{ij})(1-c_{k\ell})c_{i\ell}c_{kj} \tag{6}$$

$$I_{ijk\ell;2}(\boldsymbol{c}) = c_{ij}c_{\ell k}(1-c_{ik})(1-c_{\ell j}) + (1-c_{ij})(1-c_{\ell k})c_{ik}c_{\ell j} \tag{7}$$

$$I_{ijk\ell;3}(\boldsymbol{c}) = c_{ik}c_{j\ell}(1-c_{i\ell})(1-c_{jk}) + (1-c_{ik})(1-c_{j\ell})c_{i\ell}c_{jk} \tag{8}$$

If $I_{ijk\ell;\alpha}(\boldsymbol{c}) = 1$, this edge swap will operate as follows:

$$F_{ijk\ell;\alpha}(\boldsymbol{c})_{qr} = 1-c_{qr} \quad \text{for } (q,r) \in \mathcal{S}_{ijk\ell;\alpha} \tag{9}$$

$$F_{ijk\ell;\alpha}(\boldsymbol{c})_{qr} = c_{qr} \quad \text{for } (q,r) \notin \mathcal{S}_{ijk\ell;\alpha} \tag{10}$$

where

$$\mathcal{S}_{ijk\ell;1} = \{(i,j),(k,\ell),(i,\ell),(k,j)\} \tag{11}$$

$$\mathcal{S}_{ijk\ell;2} = \{(i,j),(\ell,k),(i,k),(\ell,j)\} \tag{12}$$

$$\mathcal{S}_{ijk\ell;3} = \{(i,k),(j,\ell),(i,\ell),(j,k)\} \tag{13}$$

Insertion of these definitions into the recipe (2,3,4) gives

$$W(\boldsymbol{c}|\boldsymbol{c}') = \sum_{i<j<k<\ell}\sum_{\alpha\leq 3}\frac{I_{ijk\ell;\alpha}(\boldsymbol{c}')}{n(\boldsymbol{c}')} \times\left[\frac{\delta_{\boldsymbol{C},F_{ijk\ell;\alpha}\boldsymbol{C}'}\mathrm{e}^{-\frac{1}{2}[E(F_{ijk\ell;\alpha}\boldsymbol{C}')-E(\boldsymbol{C}')]}+\delta_{\boldsymbol{C},\boldsymbol{C}'}\mathrm{e}^{\frac{1}{2}[E(F_{ijk\ell;\alpha}\boldsymbol{C}')-E(\boldsymbol{C}')]}}{\mathrm{e}^{-\frac{1}{2}[E(F_{ijk\ell;\alpha}\boldsymbol{C}')-E(\boldsymbol{C}')]}+\mathrm{e}^{\frac{1}{2}[E(F_{ijk\ell;\alpha}\boldsymbol{C}')-E(\boldsymbol{C}')]}}\right] \tag{14}$$

with $E(\boldsymbol{c}) = H(\boldsymbol{c})+\log n(\boldsymbol{c})$. The process (14) can be described as the following algorithm. Given an instantaneous graph $\boldsymbol{c}'$: (i) pick uniformly at random a quadruplet (i,j,k,ℓ) of sites, (ii) if at least one of the three edge swaps $\boldsymbol{c}' \to \boldsymbol{c} = F_{ijk\ell;\alpha}(\boldsymbol{c}')$ is possible, select one of these uniformly at random and execute it with an acceptance probability

$$A(\boldsymbol{c}|\boldsymbol{c}') = \left[1+\mathrm{e}^{E(F_{ijk\ell;\alpha}\boldsymbol{C}')-E(\boldsymbol{C}')}\right]^{-1} \tag{15}$$

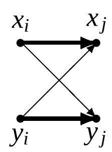
then return to (i). For this Markov chain recipe to be practical we finally need a formula for the mobility $n(\boldsymbol{c})$ of a graph. This could be calculated [9], giving[1] (with $\mathrm{Tr}\boldsymbol{A} = \sum_i A_{ii}$):

$$n(\boldsymbol{c}) = \frac{1}{4}\Big(\sum_i k_i\Big)^2 + \frac{1}{4}\sum_i k_i - \frac{1}{2}\sum_i k_i^2 - \frac{1}{2}\sum_{ij} k_i c_{ij} k_j + \frac{1}{4}\mathrm{Tr}(\boldsymbol{c}^4) + \frac{1}{2}\mathrm{Tr}(\boldsymbol{c}^3) \tag{16}$$

Naive 'accept-all' edge swapping, where $A(\boldsymbol{c}|\boldsymbol{c}') = 1$, corresponds to choosing $E(\boldsymbol{c}) = 0$ in (14), and would give the *biased* graph sampling probabilities $p_\infty(\boldsymbol{c}) = n(\boldsymbol{c})/\sum_{\boldsymbol{C}'} n(\boldsymbol{c}')$ upon equilibration. The graph mobility acts as an entropic force which can only be neglected if (16) is dominated by its first three terms; it was shown that a sufficient condition for this to be true is $\langle k^2\rangle k_{\max}/\langle k\rangle^2 \ll N$. In networks with narrow degree sequences this condition holds, and naive edge swapping is roughly acceptable. However, one has to be careful with scale-free graphs, where $\langle k^2\rangle$ and $k_{\max}$ diverge as $N \to \infty$.

IV. Degree-constrained dynamics of directed graphs

For directed networks the generalisation of the canonical edge swap (up to relabelling of nodes) is shown below:



[1]Efficient expressions for the change in mobility following one move have also been derived [9], to avoid unnecessary matrix multiplication in the computational implementation.

This move samples the space of all directed graphs with prescribed in- and out- degrees ergodically only if self-interactions are permitted [10]. We could now proceed with our formalism as before, but it is more efficient in the case of directed graphs to define the swaps and the indicator function in terms of pairs of links rather than quadruples of nodes.

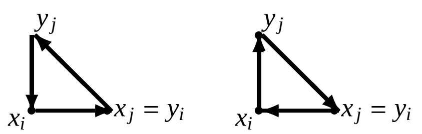
Let $\boldsymbol{c}$ now be a nonsymmetric connectivity matrix, such that $c_{ab} = 1$ if $a \to b$, otherwise $c_{ab} = 0$. Consider Λ to be the set of links within the network defined by $\boldsymbol{c}$. We write $\mathbf{x} = (x_i, x_j) \in \Lambda$ if and only if $c_{x_i x_j} = 1$. The indicator function can be written as

$$I_{\mathbf{x},\mathbf{y};\square} = \begin{cases} 1 & \text{if } \mathbf{x},\mathbf{y} \in \Lambda \text{ and } (x_i, y_j), (y_i, x_j) \notin \Lambda \\ 0 & \text{otherwise} \end{cases}$$

If $I_{\mathbf{x},\mathbf{y};\square} = 1$, then the corresponding autoinvertible operation $F_{\mathbf{x},\mathbf{y};\square}$ acts on state $\boldsymbol{c}$ as follows:

$$F_{\mathbf{x},\mathbf{y};\square}(\boldsymbol{c})_{qr} = 1 - c_{qr} \quad \text{for } q \in \{x_i, y_i\} \text{ and } r \in \{x_j, y_j\} \tag{17}$$
$$F_{\mathbf{x},\mathbf{y};\square}(\boldsymbol{c})_{qr} = c_{qr} \quad \text{otherwise} \tag{18}$$

When self-interactions are forbidden, a further type of elementary move is required to ensure ergodicity [10]. It can be visualised as reversing a triangle cycle:



The indicator function for this new move is

$$I_{\mathbf{x},\mathbf{y};\triangle} = \begin{cases} 1 & \text{if } \mathbf{x},\mathbf{y},(y_j, x_i) \in \Lambda \text{ and } \mathrm{x_j} = \mathrm{y_i} \\ & \mathbf{x}^{-1}, \mathbf{y}^{-1}, (x_i, y_j) \notin \Lambda \\ 0 & \text{otherwise} \end{cases}$$

where $\mathbf{x}^{-1} = (x_j, x_i)$ represents a link that is to be reversed. The corresponding autoinvertible operation represented by $F_{\mathbf{x},\mathbf{y};\triangle}$ acts on state $\boldsymbol{c}$ as follows:

$$F_{\mathbf{x},\mathbf{y};\triangle}(\boldsymbol{c})_{qr} = 1 - c_{qr} \quad \text{for } (q, r) \in \mathcal{S}_{x_i, x_j, y_j} \tag{19}$$
$$F_{\mathbf{x},\mathbf{y};\triangle}(\boldsymbol{c})_{qr} = c_{qr} \quad \text{for } (q, r) \notin \mathcal{S}_{x_i, x_j, y_j} \tag{20}$$

in which $\mathcal{S}_{abc} = \{(a, b), (b, c), (c, a), (b, a), (c, b), (a, c)\}$ is the set of pairs of vertices of the triangle.

We now combine edge swaps and triangular reversions into (2). This requires us to express $\sum_{F\in\Phi} q(F|\boldsymbol{c}')$ from the point of view of sampling links $\mathbf{x}$ and $\mathbf{y}$. This is in fact straightforward, as each pair of links can either trigger the square indicator function (if they define an edge swap), a triangle indicator function (if they define a triangle reversion), or neither (when the indicator function returns zero). It only remains to observe the symmetries of the problem: cycling through each pair of bonds we will come across each unique edge swap twice, and across each unique triangle swap three times. Hence

$$\sum_{F\in\Phi} q(F|\boldsymbol{c}') = \sum_{\mathbf{x},\mathbf{y}\in\Lambda} \frac{\frac{1}{2} I_{\mathbf{x},\mathbf{y};\square} + \frac{1}{3} I_{\mathbf{x},\mathbf{y};\triangle}}{n(\boldsymbol{c}')}$$

which expresses sampling over all possible moves in terms of (the more tangible) sampling over links. With the above definitions in hand, the transition probability is given by

$$\begin{aligned} W(\boldsymbol{c}|\boldsymbol{c}') &= \sum_{\mathbf{x},\mathbf{y}\in\Lambda} \frac{\frac{1}{2} I_{\mathbf{x},\mathbf{y};\square} + \frac{1}{3} I_{\mathbf{x},\mathbf{y};\triangle}}{n(\boldsymbol{c}')} \\ &\times \Big[\frac{\delta_{\boldsymbol{c},F_{\mathbf{x},\mathbf{y}}\boldsymbol{c}'} \mathrm{e}^{-\frac{1}{2}[E(F_{\mathbf{x},\mathbf{y}}\boldsymbol{c}') - E(\boldsymbol{c}')]} + \delta_{\boldsymbol{c},\boldsymbol{c}'} \mathrm{e}^{\frac{1}{2}[E(F_{\mathbf{x},\mathbf{y}}\boldsymbol{c}') - E(\boldsymbol{c}')]}}{\mathrm{e}^{-\frac{1}{2}[E(F_{\mathbf{x},\mathbf{y}}\boldsymbol{c}') - E(\boldsymbol{c}')]} + \mathrm{e}^{\frac{1}{2}[E(F_{\mathbf{x},\mathbf{y}}\boldsymbol{c}') - E(\boldsymbol{c}')]}} \Big] \end{aligned} \tag{21}$$

where $n(\boldsymbol{c}) = n_\square(\boldsymbol{c}) + n_\triangle(\boldsymbol{c})$ is the total number of valid moves that can act on the directed graph $\boldsymbol{c}$, written as the sum of the number of edge swaps and the number of triangle reversions. The mobility terms can again be calculated, giving

$$\begin{aligned} n_\square(\boldsymbol{c}) &= \frac{1}{2} Tr(\boldsymbol{c}\boldsymbol{c}^T\boldsymbol{c}\boldsymbol{c}^T) - \sum_{i,j} \mathbf{k}_i^{out} c_{ij} \mathbf{k}_j^{in} + Tr(\boldsymbol{c}\boldsymbol{c}^T\boldsymbol{c}) \\ &+ \frac{1}{2}(\sum_i \mathbf{k}_i)^2 + \frac{1}{2} Tr(\boldsymbol{c}^2) - \sum_i \mathbf{k}_j^{out} \mathbf{k}_j^{in} \\ n_\triangle(\boldsymbol{c}) &= \frac{1}{3} Tr(\boldsymbol{c}^3) - Tr(\mathbf{c}^{\updownarrow}\mathbf{c}^2) + Tr(\mathbf{c}^{\updownarrow 2}\mathbf{c}) - \frac{1}{3} Tr(\mathbf{c}^{\updownarrow 3}) \end{aligned}$$

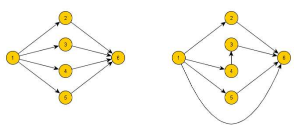
with $(\boldsymbol{c}^T)_{ij} = c_{ji}$ and $(\boldsymbol{c}^\updownarrow)_{ij} = c_{ij}c_{ji}$. Also for directed graphs it is possible to calculate efficient expressions for how the mobilities change following a single move.

For the simplest case where $H(\boldsymbol{c})$ is constant, i.e. the Markov chain is to evolve towards a flat measure, we immediately observe that the required move acceptance probabilities are

$$A(\boldsymbol{c}|\boldsymbol{c}') = \Big[1 + \frac{n_\square(\boldsymbol{c}) + n_\triangle(\boldsymbol{c})}{n_\square(\boldsymbol{c}') + n_\triangle(\boldsymbol{c}')}\Big]^{-1} \tag{22}$$

V. Numerical examples for directed graphs

Examples illustrating the canonical Markov chains for nondirected graphs are given in [9]. Here we compare two variants of the *directed* edge rewiring algorithm: the naive 'accept all moves' version, and the version with the canonical mobility corrections (22). Consider a network with $K + 2$ nodes, of which one has degrees $(k^{\text{in}}, k^{\text{out}}) = (0, K)$, one has $(k^{\text{in}}, k^{\text{out}}) = (K, 0)$, and the remaining K have degrees $(k^{\text{in}}, k^{\text{out}}) = (1, 1)$. Self-interactions are forbidden. Up to relabelling, there are two such graph types (see below):



K=25	accept all process avg. $n(\boldsymbol{c})$	correct process avg. $n(\boldsymbol{c})$
predict	58.52	47.92
actual	58.32	47.95

The left graph type has mobility $K(K-1)$, and only occurs in one network configuration. The right graph type has mobility of $2K-3$, and multiplicity $K(K-1)$. The table shows the results of numerical simulations for $K = 25$, compared to theoretical predictions. We have used the mobility itself as a marker for the proportion of time spent in each type of configuration; the difference between the two processes is striking, and will translate into differences in measurements of any observable which differentiates between the two configurations.

VI. Generation of random graphs with prescribed degree correlations via rewiring algorithms

Our approach can be extended to accurately target desired degree correlations defined by $W(k,k')$. This can be achieved for undirected graphs by ensuring convergence of the above Markov chain to the following non-uniform measures [11],

$$p(\boldsymbol{c}) = \frac{\delta_{\boldsymbol{k},\boldsymbol{k}(\boldsymbol{c})}}{Z}\prod_{i<j}\left[\frac{\langle k\rangle}{N}\frac{W(k_i,k_j)}{p(k_i)p(k_j)}\delta_{c_{ij},1} + \left(1-\frac{\langle k\rangle}{N}\frac{W(k_i,k_j)}{p(k_i)p(k_j)}\right)\delta_{c_{ij},0}\right] \tag{23}$$

with $p(k) = N^{-1}\sum_i \delta_{k_i,k}$, $W(k,k') = (N\langle k\rangle)^{-1}\sum_{ij} c_{ij}\delta_{k_i,k}\delta_{k_j,k'}$, $\langle k\rangle = \sum_k kp(k)$, and

$$W(k) = \sum_{k'} W(k,k') = p(k)k/\langle k\rangle \tag{24}$$

In the language of our process, (23) corresponds to the choice

$$H(\boldsymbol{c}) = -\sum_{i<j}\log\left[\frac{\langle k\rangle}{N}\frac{W(k_i,k_j)}{p(k_i)p(k_j)}\delta_{c_{ij},1} + \left(1-\frac{\langle k\rangle}{N}\frac{W(k_i,k_j)}{p(k_i)p(k_j)}\right)\delta_{c_{ij},0}\right] \tag{25}$$

Hence, for the candidate edge swaps $\boldsymbol{c}' \to \boldsymbol{c} = F_{ijk\ell;\alpha}\boldsymbol{c}'$ the acceptance probability (4) can be used, where

$$e^{H(\boldsymbol{C})-H(\boldsymbol{C}')} = \prod_{(a,b)\in S_{ijk\ell;\alpha}}\left[L_{ab}\delta_{c_{ab},1} + L_{ab}^{-1}\delta_{c_{ab},0}\right] \tag{26}$$

with $L_{ab} = N\overline{k}/[\Pi(k_a,k_b)k_ak_b] - 1$ and relative degree correlations $\Pi(k,k') = W(k,k')/W(k)W(k')$.

We show an example of the Markov chain (14) targeting (23) where relative degree correlations are chosen as

$$\Pi(k,k') = \frac{(k-k')^2}{[\beta_1-\beta_2k+\beta_3k^2][\beta_1-\beta_2k'+\beta_3k'^2]} \tag{27}$$

(the parameters β_i follow from (24)). An initial graph $\boldsymbol{c}_0$ was constructed with a non-Poissonian degree distribution and no degree correlations, i.e. $\Pi(k,k'|\boldsymbol{c}_0) \approx 1$. After iterating the Markov chain until equilibrium (after $75,000$ accepted moves, and after reaching maximal Hamming distance between initial and final configuration) degree correlations are seen in very good agreement with their target values; (see Figure 1). Extension to directed graphs is achieved by replacing k with $\boldsymbol{k} = (k^{\text{in}}, k^{\text{out}})$, allowing repetition of site indices in (26) and bearing in mind that $\Pi(\boldsymbol{k}_a,\boldsymbol{k}_b) \neq \Pi(\boldsymbol{k}_b,\boldsymbol{k}_a)$.

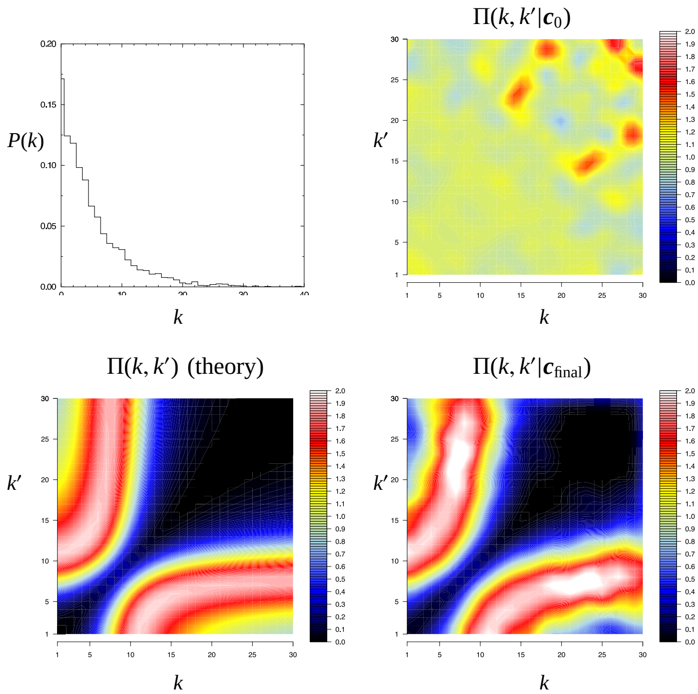


Fig. 1. Results of Markovian dynamics tailored to target the non-uniform measure (23). Top left: degree distribution of the (randomly generated) initial graph $\boldsymbol{c}_0$, with $N = 4000$ and $\langle k\rangle = 5$. Top right: $\Pi(k,k'|\boldsymbol{c}_0)$ of the initial graph. Bottom left: the target relative degree correlations (27) chosen in (23). Bottom right: colour plot of $\Pi(k,k'|\boldsymbol{c}_{\text{final}})$ in the final graph $\boldsymbol{c}_{\text{final}}$, measured after 75,000 accepted moves of the Markov chain (14)

VII. Discussion

In this paper we focused on how to generate numerically tailored random graphs with controlled macroscopic structural properties, to serve e.g. as null models in hypothesis testing. Bias in the generation of random graphs has the potential to invalidate all further statistical analysis performed on the generated networks and has been well documented in the literature; it is known to affect the 'stubs' method and the 'accept-all' edge swapping method. However, the lack so far of workable corrections or alternatives has meant that these issues have often been ignored. Our theory offers a practical and theoretically sound approach to uniformly generating random graphs from ensembles which share certain topological characteristics with a real network, and can therefore serve as a reliable tool for building unbiased null models.

References

[1] E. Bender and E. Canfield, "The asymptotic of labelled graphs with given degree sequences," *J. Comb. Theory, Ser. A*, vol. 24, pp. 296–307, 1978.

[2] M. Molloy and B. Reed, "A critical point for random graphs with a given degree sequence," *Random Structures and Algorithms*, vol. 6, pp. 161–180, 1995.

[3] F. Chung and L. Lu, "The average distances in random graphs with given expected degrees," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 99, pp. 15 879–15 882, 2002.

[4] A. Stauffer and V. Barbosa, "A study of the edge switching markov-chain method for the generation of random graphs," *preprint arXiv:0512105*, 2005.

[5] M. Newman, S. Strogatz, and D. Watts, *Phys. Rev. E*, vol. 64, 2001.

[6] H. Klein-Henning and A. K. Hartman, "Bias in generation of random graphs," *preprint arXiv:1107.5734*, 2011.

[7] R. B. Eggleton and D. A. Holton, M. KL, Ed. Springer Lect. Notes Math., 1981, no. 884.

[8] R. Taylor, *Constrained switchings in graphs.*, M. KL, Ed. Springer Lect. Notes Math., 1981, no. 884.

[9] A. C. C. Coolen, A. De Martino, and A. Annibale, "Constrained Markovian dynamics of random graphs," *J. Stat. Phys*, vol. 136, no. 6, pp. 1035–1067, 2009.

[10] A. R. Rao, R. Jana, and S. Bandyopadhyay, "A Markov Chain Monte Carlo Method for Generating Random (0, 1)-Matrices with Given Marginals," *Sankhy: The Indian Journal of Statistics, Series A*, vol. 58, no. 2, 1996.

[11] C. Perez-Vicente and A. Coolen, "Spin models on random graphs with controlled topologies beyond degree constraints," *J. Phys. A*, vol. 41, p. 255003, 2008.