Tailoring of Random Networks and Graphs
ACC Coolen, King’s College London

Eindhoven, Sept 21st 2017
Introduction
  Networks and graphs
  Tailored random graph ensembles

Counting tailored random graphs
  Entropy and complexity
  Nondirected graphs
  Directed graphs

Generating tailored random graphs
  Common algorithms and their problems
  MCMC processes for hard-constrained graphs

Degree-constrained MCMC graph dynamics
  Bookkeeping of moves
  Mobility of nondirected graphs
  Directed graphs

Tailoring loopy graph ensembles
  Motivation
  Spectrally constrained ensembles
  Solvable toy model
Introduction

Networks and graphs
Tailored random graph ensembles

Counting tailored random graphs
Entrophy and complexity
Nondirected graphs
Directed graphs

Generating tailored random graphs
Common algorithms and their problems
MCMC processes for hard-constrained graphs

Degree-constrained MCMC graph dynamics
Bookkeeping of moves
Mobility of nondirected graphs
Directed graphs

Tailoring loopy graph ensembles
Motivation
Spectrally constrained ensembles
Solvable toy model
Networks and graphs

nodes (vertices): \( i, j \in \{1, \ldots, N\} \)

links (edges): \( c_{ij} \in \{0, 1\} \)

no self-links: \( c_{ii} = 0 \) for all \( i \)

graph: \( c = \{c_{ij}\} \)

nondirected: \( \forall (i, j) : c_{ij} = c_{ji} \)

directed: \( \exists (i, j) : c_{ij} \neq c_{ji} \)

if we model real-world systems by random graphs
we want these graphs to be realistic ...

i.e. to have appropriate domain-specific statistical characteristics
Quantify topology of nondirected graphs

- degrees,
  degree sequence: \( k_i(c) = \sum_j c_{ij}, \quad k(c) = (k_1(c), \ldots, k_N(c)) \)

degree distribution: \( p(k|c) = \frac{1}{N} \sum_{i=1}^{N} \delta_{k_i, k} \)

- joint degree statistics of connected nodes

\[
W(k, k'|c) = \frac{1}{N\langle k \rangle} \sum_{ij} c_{ij} \delta_{k_i, k} \delta_{k_j, k'}
\]

normalisation: \( \sum_{k, k' \geq 0} W(k, k'|c) = 1 \)

assortativity/dissortativity: \( C = \langle kk' \rangle_w - \langle k \rangle_w \langle k' \rangle_w \)
marginals of $W$ carry no info beyond degree statistics,

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

so focus on:

$$\Pi(k, k'|c) = \frac{W(k, k'|c)}{W(k|c)W(k'|c)}$$

if $\exists (k, k')$ with $\Pi(k, k'|c) \neq 1$:

structural information in degree correlations

human PIN
$N = 9306$
$\langle k \rangle = 7.53$
Quantify topology of directed graphs

links become *arrows*

- degrees,
- degree sequences:
  \[ k_i^{\text{in}}(\mathbf{c}) = \sum_j c_{ij}, \quad k^{\text{in}}(\mathbf{c}) = (k_1^{\text{in}}(\mathbf{c}), \ldots, k_N^{\text{in}}(\mathbf{c})) \]
  \[ k_i^{\text{out}}(\mathbf{c}) = \sum_j c_{ji}, \quad k^{\text{out}}(\mathbf{c}) = (k_1^{\text{out}}(\mathbf{c}), \ldots, k_N^{\text{out}}(\mathbf{c})) \]

degree distribution:

\[ k_i \rightarrow \vec{k}_i = (k_i^{\text{in}}, k_i^{\text{out}}) \quad p(\vec{k} | \mathbf{c}) = \frac{1}{N} \sum_i \delta_{\vec{k}, \vec{k}_i(\mathbf{c})} \]

- joint in-out degree statistics
  of connected nodes

\[ W(\vec{k}, \vec{k}' | \mathbf{c}) = \frac{1}{N \langle k \rangle} \sum_{ij} c_{ij} \delta_{\vec{k}, \vec{k}_i(\mathbf{c})} \delta_{\vec{k}', \vec{k}_j(\mathbf{c})} \]

note:

\[ W(\vec{k}, \vec{k}' | \mathbf{c}) \neq W(\vec{k}', \vec{k} | \mathbf{c}) \]
Graph classification via increasingly detailed feature prescription

$G = \{0, 1\}_\frac{1}{2}N(N-1)$

$\langle k \rangle = \ldots$

$p(k) = \ldots$

$W(k, k') = \ldots$

Tailoring random graphs

maximum entropy random graph ensembles, $p(c)$ with prescribed values for $\langle k \rangle, p(k), W(k, k'), \ldots$

– proxies for real networks in stat mech models
– complexity: how many graphs have same features as $c$? counting
– hypothesis testing: graphs as null models generation

$N = 1000$: $2^{\frac{1}{2}N(N-1)} \approx 10^{150,364}$ graphs
(universe has $\sim 10^{82}$ atoms ...)
Tailored random graph ensembles

(i) set $G$ of allowed graphs,
(ii) probability measure $p(c)$ on $G$

▶ Tailoring via hard constraints

impose values for observables: $\Omega_\mu(c) = \Omega_\mu$ for $\mu = 1 \ldots p$

$$p(c|\Omega) = \frac{\delta_{\Omega(c),\Omega}}{\mathcal{N}(\Omega)}, \quad \mathcal{N}(\Omega) = \sum_c \delta_{\Omega(c),\Omega} \quad (# \text{ graphs in ensemble})$$

note:

maximises Shannon entropy $S$
on $G[\Omega] = \{c \mid \Omega(c) = \Omega\}$

$$S = -\frac{1}{N\langle k \rangle} \sum_c p(c) \log p(c)$$

$$e^{N\langle k \rangle S[\Omega]} = e^{-\sum_c \frac{\delta_{\Omega(c),\Omega}}{\mathcal{N}(\Omega)} \left( \log \delta_{\Omega(c),\Omega} - \log \mathcal{N}(\Omega) \right)} = \mathcal{N}(\Omega)$$

with $\Omega = (\Omega_1, \ldots, \Omega_p)$
Tailoring via soft constraints

impose averages for observables: \( \Omega_\mu(c) = \Omega_\mu \) for \( \mu = 1 \ldots p \)

\( p(c) \): maximum entropy, subject to constraints

\[
p(c|\Omega) = Z^{-1}(\Omega) e^{\sum_\mu \omega_\mu \Omega_\mu(c)}
\]

parameters \( \omega_\mu \):

to be solved from

\[
\forall \mu : \sum_c p(c|\Omega) \Omega_\mu(c) = \Omega_\mu
\]

now all graphs \( c \) can emerge,

but those with \( \Omega(c) \approx \Omega \) are most likely

effective # graphs \( N(\Omega) \) defined via entropy:

\[
N(\Omega) = e^{N\langle k \rangle S[\Omega]}, \quad S[\Omega] = -\frac{1}{N\langle k \rangle} \sum_{c \in G} p(c|\Omega) \log p(c|\Omega)
\]
Example

nondirected graphs, $c_{ii} = 0$ for all $i$, impose average connectivity via hard constraint, $\Omega(c) = \sum_{ij} c_{ij}$

- demand $\sum_{ij} c_{ij} = N \langle k \rangle$

$$p(c|\langle k \rangle) = \frac{\delta_{\sum_{ij} c_{ij}, N\langle k \rangle}}{\mathcal{N}(\langle k \rangle)}, \quad \mathcal{N}(\langle k \rangle) = \sum_{c} \delta_{\sum_{ij} c_{ij}, N\langle k \rangle}$$

- calculate $\mathcal{N}(\langle k \rangle)$:
  
  use $\delta_{nm} = (2\pi)^{-1} \int_{-\pi}^{\pi} d\omega \ e^{i(n-m)\omega}$

$$\mathcal{N}(\langle k \rangle) = \left( \frac{1}{2} \frac{N(N-1)}{N\langle k \rangle} \right)$$

$$= e^{\frac{1}{2} N\langle k \rangle \left[ \log(N/\langle k \rangle) + 1 \right]} + \mathcal{O}(\log N)$$
Example

nondirected graphs, $c_{ii} = 0$ for all $i$,
impose average connectivity via soft constraint,$\Omega(c) = \sum_{ij} c_{ij}$

- demand $\langle \sum_{ij} c_{ij} \rangle = N\langle k \rangle$
  \[ p(c|\langle k \rangle) = \frac{1}{Z(\omega)} e^{\omega \sum_{ij} c_{ij}}, \quad Z(\omega) = \sum_c e^{\omega \sum_{ij} c_{ij}} \]

$\omega$ solved from:  $\langle k \rangle = \frac{d}{d\omega} \frac{1}{N} \log Z(\omega)$

- calculate $Z(\omega)$ and $\omega$:
  \[ \langle k \rangle = (N-1) \frac{e^{2\omega}}{e^{2\omega} + 1} \]

- rewrite probabilities:
  \[ p(c|\langle k \rangle) = \prod_{i<j} \left[ \frac{e^{2\omega}}{e^{2\omega} + 1} \delta_{c_{ij},1} + \frac{1}{e^{2\omega} + 1} \delta_{c_{ij},0} \right] \]

Erdős-Rényi ensemble
Introduction
   Networks and graphs
   Tailored random graph ensembles

Counting tailored random graphs
   Entropy and complexity
   Nondirected graphs
   Directed graphs

Generating tailored random graphs
   Common algorithms and their problems
   MCMC processes for hard-constrained graphs

Degree-constrained MCMC graph dynamics
   Bookkeeping of moves
   Mobility of nondirected graphs
   Directed graphs

Tailoring loopy graph ensembles
   Motivation
   Spectrally constrained ensembles
   Solvable toy model
Counting tailored random graphs

how many graphs in each family?

\[ G = \{0, 1\}^{\frac{1}{2}N(N-1)} \]

\[ \langle k \rangle = \ldots \]

\[ p(k) = \ldots \]

\[ W(k, k') = \ldots \]

Note:

solving models of *interacting particle systems* on tailored random graphs (via replica method or generating functional analysis, for \( N \to \infty \)):
feasible if we can compute the entropy of the graph ensemble!
entropy and complexity

- **effective nr of graphs** in ensemble $p(c|\Omega)$:
  ($\Omega$: values of imposed observables)

  \[ \mathcal{N}(\Omega) = e^{N\langle k \rangle S(\Omega)}, \quad S(\Omega) = -\frac{1}{N\langle k \rangle} \sum_{c \in G} p(c|\Omega) \log p(c|\Omega) \]

- $S(\Omega)$: proportional to average nr of bits we need to specify to identify a graph $c$ in the ensemble

- **complexity of graphs** in ensemble $p(c|\Omega)$:

  \[ C(\Omega) = S(\emptyset) - S(\Omega) \]

  $\emptyset$: no constraints
  nondirected, $c_{ii} = 0 \forall i$:

  \[ p(c|\emptyset) = 2^{-\frac{1}{2}N(N-1)}, \quad S(\emptyset) = -\frac{1}{N\langle k \rangle} \log 2^{-\frac{1}{2}N(N-1)} = \frac{N-1}{2\langle k \rangle} \log 2 \]
Nondirected graphs

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

\[
S = \frac{1}{2} \left[ 1 + \log \left( \frac{N}{\langle k \rangle} \right) \right] - \left\{ \frac{1}{\langle k \rangle} \sum_k p(k) \log \left[ \frac{p(k)}{\tilde{p}(k)} \right] + \frac{1}{2} \sum_{k, k'} W(k, k') \log \left[ \frac{W(k, k')}{W(k)W(k')} \right] \right\}
\]

\[
\lim_{N \to \infty} \epsilon_N = 0
\]

\[
\tilde{p}(\ell) = e^{-\langle k \rangle \langle k \rangle^{\ell}} / \ell!
\]

degree distr of Erdős-Renyi graphs

(path integrals, integral representations, steepest descent, ...)

"Erdos-Renyi entropy"

"degree complexity"

"wiring complexity"

\[
S = \frac{1}{2} \left[ 1 + \log \left( \frac{N}{\langle k \rangle} \right) \right] - \left\{ \frac{1}{\langle k \rangle} \sum_k p(k) \log \left[ \frac{p(k)}{\tilde{p}(k)} \right] + \frac{1}{2} \sum_{k, k'} W(k, k') \log \left[ \frac{W(k, k')}{W(k)W(k')} \right] \right\}
\]

\[
\lim_{N \to \infty} \epsilon_N = 0
\]

\[
\tilde{p}(\ell) = e^{-\langle k \rangle \langle k \rangle^{\ell}} / \ell!
\]

degree distr of Erdős-Renyi graphs

(path integrals, integral representations, steepest descent, ...)
Directed graphs

\( \vec{k}_i = (k_i^\text{in}, k_i^\text{out}) \)

\[
p(c) = \sum \prod_i \left[ d\vec{k}_i \ p(\vec{k}_i) \right] \frac{\prod_i \delta_{\vec{k}_i, \vec{k}_i(c)}}{Z(k, W)} \prod_{i<j} \left[ \frac{\langle k \rangle \ W(\vec{k}_i, \vec{k}_j)}{\langle k \rangle \ p(\vec{k}_i)p(\vec{k}_j)} \delta_{c_{ij}, 1} + \left( 1 - \frac{\langle k \rangle \ W(\vec{k}_i, \vec{k}_j)}{\langle k \rangle \ p(\vec{k}_i)p(\vec{k}_j)} \right) \delta_{c_{ij}, 0} \right]
\]

\[
S = 1 + \log\left( \frac{N}{\langle k \rangle} \right) - \left\{ \frac{1}{\langle k \rangle} \sum \ p(\vec{k}) \log\left[ \frac{p(\vec{k})}{\tilde{p}(k^\text{in})\tilde{p}(k^\text{out})} \right] + \sum \ W(\vec{k}, \vec{k}') \log\left[ \frac{W(\vec{k}, \vec{k}')}{W(\vec{k})W(\vec{k}')} \right] \right\}
\]

\[
\lim_{N \to \infty} \epsilon_N = 0
\]

\( \tilde{p}(\ell) = e^{-\langle k \rangle \langle k \rangle \ell / \ell!} \)

\( \tilde{p}(k^\text{in})\tilde{p}(k^\text{out}) \): degree distr of directed Erdős-Renyi graphs

(path integrals, integral representations, steepest descent, ...)

\( \epsilon_N \)
Introduction
 Networks and graphs
 Tailored random graph ensembles

Counting tailored random graphs
 Entropy and complexity
 Nondirected graphs
 Directed graphs

Generating tailored random graphs
 Common algorithms and their problems
 MCMC processes for hard-constrained graphs

Degree-constrained MCMC graph dynamics
 Bookkeeping of moves
 Mobility of nondirected graphs
 Directed graphs

Tailoring loopy graph ensembles
 Motivation
 Spectrally constrained ensembles
 Solvable toy model
Generating tailored random graphs

\[ G = \{0, 1\}^{\frac{1}{2}N(N-1)} \]

\[ \langle k \rangle = \ldots \]

\[ k = \ldots \]

\[ W(k, k') = \ldots \]

\( G \): all nondirected \( N \)-node graphs

\( G[k] \subset G \): all nondirected \( N \)-node graphs with degrees \( k \)

typical questions:
how to generate numerically

- random \( c \in G \), with specified \( p(c) \)
- random \( c \in G[k] \), with uniform \( p(c) \)
- random \( c \in G[k] \), with specified \( p(c) \)

similar for directed graphs ...
Common algorithms and their problems

soft constraints only:
standard Glauber/Gibbs/MCMC dynamics

objective: generate random nondirected \( c \in \{0, 1\}^\frac{1}{2}N(N-1) \)
with specified probabilities \( p(c) \)

strategy: start from any graph \( c \)
propose random moves \( c_{ij} \rightarrow 1 - c_{ij} \) (giving \( c \rightarrow F_{ij}c \)),
define acceptance probabilities \( A(F_{ij}c|c) \)
via detailed balance condition

\[
A(F_{ij}c|c)p(c) = A(c|F_{ij}c)p(F_{ij}c) \quad \rightarrow \quad A(c'|c) = \left[ 1 + p(c)/p(c') \right]^{-1}
\]

stochastic process is ergodic, and converges to \( p(c) \)

practicalities:
equilibration can take a very long time,
so monitor Hamming distances similar for directed graphs ...
The problem of phase transitions

example:

\[ p(c) = \frac{1}{Z(\alpha, \beta)} e^{\alpha \sum_i k_i(c) + \beta \sum_i k_i^2(c)}, \quad N=300, \quad \alpha = 4 \]

- phase transitions sometimes prevent us from controlling observables in soft-constrained ensembles
- need hard constrained ensembles ...
  but these are harder to sample via MCMC ...
Matching algorithm  
(Bender and Canfield, 1978)

objective: generate random nondirected graph \( c \in \{0, 1\}^{N(N-1)/2} \) with specified degree sequence \( k = (k_1, \ldots, k_N) \)

strategy: stochastic growth dynamics, starting from graph with no links

- initialisation: \( c_{ij} = 0 \) for all \((i, j)\)

 repeat:

 - pick at random two nodes \((i, j)\)
 - if \( \sum_{\ell} c_{i\ell} < k_i \) and \( \sum_{\ell} c_{j\ell} < k_j \):
   connect \(i\) and \(j\)
   \( c_{ij} = 0 \rightarrow c_{ij} = 1 \)

 terminate if \( \sum_j c_{ij} = k_i \) for all \(i\)

(trivially generalised to directed graphs)
Matching algorithm limitations and problems ...

- major limitation:
  
  aims to generate random \( c \in G[k] \), but cannot control graph probabilities ...

- inconvenience: convergence not guaranteed

  process can ‘hang’ before \( \sum_j c_{ij} = k_i \) for all \( i \)
  if a remaining ‘stub’ requires self-loops
  – monitor evolving degrees, to test for this
  – if process ‘hangs’: reject and start again from empty graph

- sampling bias:

  if process ‘hangs’, users often don’t reject the graph but do ‘backtracking’ (for CPU reasons),
  this creates correlations between graph realisations
  even if we reject rather than backtrack:
  no proof published yet that sampling measure \( p(c) \) is flat ...
MCMC with hard constraints

need to think more carefully about elementary moves in space of graphs

<table>
<thead>
<tr>
<th>MOVE SET</th>
<th>INVARIANTS</th>
<th>ACTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Link flips</td>
<td>none</td>
<td><img src="image" alt="Link flip" /></td>
</tr>
<tr>
<td>{F_{ij}}</td>
<td></td>
<td><img src="image" alt="Link flip" /></td>
</tr>
<tr>
<td>Hinge flips</td>
<td>average degree (\bar{k}(c) = \frac{1}{N} \sum_{rs} c_{rs})</td>
<td><img src="image" alt="Hinge flip" /></td>
</tr>
<tr>
<td>{F_{ijk}}</td>
<td></td>
<td><img src="image" alt="Hinge flip" /></td>
</tr>
<tr>
<td>Edge swaps</td>
<td>all individual degrees (k_i(c) = \sum_j c_{ij}, \ i = 1 \ldots N)</td>
<td><img src="image" alt="Edge swap" /></td>
</tr>
<tr>
<td>{F_{ijk\ell}}</td>
<td></td>
<td><img src="image" alt="Edge swap" /></td>
</tr>
</tbody>
</table>
Edge switching algorithm
(Seidel, 1976)

objective: generate random nondirected graph $\mathbf{c} \in \{0, 1\}^{1/2N(N-1)}$
with specified degree sequence $\mathbf{k} = (k_1, \ldots, k_N)$

strategy: degree-preserving randomisation (‘shuffling’) process,
starting from any graph $\mathbf{k} = (k_1, \ldots, k_N)$

- initialisation: $c_{ij} = c_{ij}^0$ for all $(i, j)$,
  $\mathbf{c}^0$: any graph
  with the correct degrees

repeat:
- pick at random four nodes $(i, j, k, \ell)$
  that are pairwise connected
- carry out an ‘edge swap’
  (or ‘Seidel switch’), see diagram
  (preserves all degrees!)

terminate if stochastic process has equilibrated
Edge switching algorithm
limitations and problems ...

- major limitation:
  aims to generate random \( \mathbf{c} \in G[k] \), but cannot control graph probabilities ...

- inconvenience: need for a ‘seed graph’ with the correct degrees \( k = (k_1, \ldots, k_N) \)

- sampling bias:
  edge swaps are ergodic on \( G[k] \) (Taylor, 1981), but sampling is not uniform!

\[
\text{many possible moves} \quad \text{few moves} ...
\]

nr of possible moves depends on state \( \mathbf{c} \)!

result:
stationary state of Markov chain favours high-mobility graphs

dangerous for scale-free graphs ...
target: uniform measure $p(c)$ on $G[k]$

$n(c)$ : nr of possible moves

for flat measure:

$$\overline{n(c)} = \frac{(N-2)(N-3)[1 + 2(N-3)]}{1 + (N-2)(N-3)}$$

$N = 100$:

$$\overline{n(c)}/N^2 \approx 0.0195$$

‘accept all’ edge swapping:

$$\overline{n(c)}/N^2$$
why is the generation of graphs with hard constraints nontrivial?

- many users underestimate/misjudge what the real problem is: sampling space of all graphs with given features: usually easy ... sampling them with specified probabilities: nontrivial!

- many ad-hoc graph generation algorithms appear sensible, but lack analysis of which measure they converge to

random graphs are often used as ‘null models’, against which to test hypotheses on real networks.

if null model is biased, hypothesis test is fundamentally flawed ...
MCMC processes for hard-constrained graphs

- **hard constraints:**
  \[ G[\star] \subseteq G: \quad \forall c \in G \text{ that satisfy constraints } \star \]

- **stochastic graph dynamics as a Markov chain,**
  transition probabilities \( W(c|c') \) for move \( c' \rightarrow c \)

\[
\forall c \in G[\star]: \quad \rho_{t+1}(c) = \sum_{c' \in G[\star]} W(c|c') \rho_t(c')
\]

- **allowed moves (exclude identity):**
  \( \Phi \): set of allowed moves \( F: G_F[\star] \rightarrow G[\star] \)
  \( G_F[\star] \): those \( c \in G[\star] \) on which \( F \) can act
  all moves are auto-invertible: \( \forall F \in \Phi \): \( F^2 = I \)
  \( \Phi \) is ergodic on \( G[\star] \)

**objective**

construct transition probs on \( G[\star] \), based on move set \( \Phi \),
such that process converges to \( p(c) = Z^{-1} e^{-H(c)} \)
standard form:

\[ W(c|c') = \sum_{F \in \Phi} q(F|c') \left[ \delta_{c,Fc'} A(Fc'|c') + \delta_{c,c'} [1 - A(Fc'|c')] \right] \]

- \( q(F|c) \): move proposal probability
- \( A(c|c') \): move acceptance probability

detailed balance:

\[ (\forall F \in \Phi)(\forall c \in G[*]) : \quad q(F|c)A(Fc|c)e^{-H(c)} = q(F|Fc)A(c|Fc)e^{-H(Fc)} \]

move proposal probability:

\[ q(F|c) = \begin{cases} 
0 & \text{if } F \text{ cannot act on } c \\
1/n(c) & \text{if } F \text{ can act on } c 
\end{cases} \]

graph mobility \( n(c) \):

\[ n(c) = \sum_{F \in \Phi} l_F(c), \quad l_F(c) = \begin{cases} 
1 & \text{if } c \in G_F[*] \\
0 & \text{if } c \notin G_F[*] 
\end{cases} \]
canonical Markov chain

ergodic auto-invertible moves $F \in \Phi$, convergence to $p(c) = Z^{-1} e^{-H(c)}$ on $G[*]$
for acceptance probabilities

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

naive edge-swapping?
$(\forall c, c') : A(c|c') = 1$

$$(\forall F, c) : \frac{A(Fc|c)e^{-H(c)}}{n(c)} = \frac{A(c|Fc)e^{-H(Fc)}}{n(Fc)} \quad \rightarrow \quad (\forall F, c) : \frac{e^{-H(c)}}{n(c)} = \frac{e^{-H(Fc)}}{n(Fc)}$$

corresponds to
$H(c) = - \log n(c)$,
so would give

$sampling bias : \quad p(c) = \frac{n(c)}{\sum_{c' \in G[*]} n(c')} $
picking moves randomly ...

correct sampling: \( q(F|c) = 1/n(c) \)
for all possible moves

---

**PROTOCOL 1:**
(i) pick a site \( j \) with \( k_j(A) > 0 \)
(ii) pick a site \( i \in \partial_j(A) \)
(iii) pick a site \( k \notin \partial_i(A) \cup \{i\} \)

---

**PROTOCOL 2:**
(i) pick two disconnected sites \( (i,k) \) with \( k_i(A) > 0 \)
(ii) pick a site \( j \in \partial_i(A) \)

---

**PROTOCOL 3:**
(i) pick two connected sites \( (i,j) \) and a third site \( k \)
(ii) while \( A_{ik} = 1 \) return to (i)
$N = 3000, \langle k \rangle = 7$

dashed: start graph

dotted: $p(k)$ of target $p(A)$

solid: MCMC result
Introduction
  Networks and graphs
  Tailored random graph ensembles

Counting tailored random graphs
  Entropy and complexity
  Nondirected graphs
  Directed graphs

Generating tailored random graphs
  Common algorithms and their problems
  MCMC processes for hard-constrained graphs

Degree-constrained MCMC graph dynamics
  Bookkeeping of moves
  Mobility of nondirected graphs
  Directed graphs

Tailoring loopy graph ensembles
  Motivation
  Spectrally constrained ensembles
  Solvable toy model
Bookkeeping of moves

- constraints: imposed degrees $k$

ergodic set $\Phi$ of admissible moves:

edge swaps $F : G_F[k] \rightarrow G[k]$

$\{(i, j, k, \ell) \in \{1, \ldots, N\}^4 | i < j < k < \ell\}$, ordered node quadruplets

- group into pairs (I,IV), (II,V), and (III,VI)

auto-invertible swaps: $F_{ijk\ell;\alpha}$, with $i < j < k < \ell$ and $\alpha \in \{1, 2, 3\}$

$I_{ijk\ell;\alpha}(c) = 1$: $F_{ijk\ell;\alpha}(c)_{qr} = 1 - c_{qr}$ for $(q, r) \in S_{ijk\ell;\alpha}$

$F_{ijk\ell;\alpha}(c)_{qr} = c_{qr}$ for $(q, r) \notin S_{ijk\ell;\alpha}$

$S_{ijk\ell;1} = \{(i, j), (k, \ell), (i, \ell), (j, k)\}$, $S_{ijk\ell;2} = \{(i, j), (k, \ell), (i, k), (j, \ell)\}$

$S_{ijk\ell;3} = \{(i, k), (j, \ell), (i, \ell), (j, k)\}$
Mobility of nondirected graphs

to implement the Markov chain, need analytical formula for the graph mobility

\[ n(c) = \sum_{i<j<k<\ell}^{N} \sum_{\alpha=1}^{3} l_{ijk\ell;\alpha}(c) \]

\[
l_{ijk\ell;1}(c) = c_{ij} c_{k\ell} (1 - c_{i\ell}) (1 - c_{jk}) + (1 - c_{ij}) (1 - c_{k\ell}) c_{i\ell} c_{jk}
\]

\[
l_{ijk\ell;2}(c) = c_{ij} c_{k\ell} (1 - c_{ik}) (1 - c_{j\ell}) + (1 - c_{ij}) (1 - c_{k\ell}) c_{ik} c_{j\ell}
\]

\[
l_{ijk\ell;3}(c) = c_{ik} c_{j\ell} (1 - c_{i\ell}) (1 - c_{jk}) + (1 - c_{ik}) (1 - c_{j\ell}) c_{i\ell} c_{jk}
\]

work out combinatorics:

\[
n(c) = \frac{1}{4} N^2 \langle k \rangle^2 + \frac{1}{4} N \langle k \rangle - \frac{1}{2} N \langle k^2 \rangle + \frac{1}{4} \text{Tr}(c^4) + \frac{1}{2} \text{Tr}(c^3) - \frac{1}{2} \sum_{ij} k_i c_{ij} k_j
\]

\[
\text{invariant}
\]

\[
\text{state dependent}
\]

- state-dependent part can be ignored if \( \langle k^2 \rangle k_{\text{max}} / \langle k \rangle^2 \ll N \)
- avoid calculating \( n(c) \) at each iteration step:
  (i) calculate \( n(c) \) at time \( t = 0 \)
  (ii) update dynamically, compute \( \Delta_{ijk\ell;\alpha} n(c) \) for executed move \( F_{ijk\ell;\alpha} \)
Example:

target = uniform measure on $G[k]$

$N = 100$

naive versus correct acceptance probabilities

predictions:

$p(c) = \text{constant}:$

$\frac{n(c)}{N^2} \approx 0.0195$

$p(c) = \frac{n(c)}{Z}:$

$\frac{n(c)}{N^2} \approx 0.0242$

\[ A(c|c') = 1 \]

\[ A(c|c') = \left[1 + \frac{n(c)}{n(c')}\right]^{-1} \]
Example

degree-correlated measure on $G[k]$

$N = 4000, \langle k \rangle = 5$

$\Pi(k, k') = \frac{(k - k')^2}{[\beta_1 - \beta_2 k + \beta_3 k^2][\beta_1 - \beta_2 k' + \beta_3 k'^2]}$
Directed graphs

bookkeeping of elementary moves

- constraints: imposed in-out degrees, so graph set is $G[k_{in}, k_{out}]$

set $\Phi$ of admissible moves:
directed edge swaps $F : G_F[k_{in}, k_{out}] \rightarrow G[k_{in}, k_{out}]$

for non-directed graphs:
edge swaps are *ergodic* set of moves
(Taylor, 1981 – proof based on Lyapunov function)

Rao, 1996:
unless self-interactions are allowed,
edge swaps *not ergodic for directed graphs*
further move type required to restore ergodicity:
3-loop reversal

to implement the Markov chain, need to calculate graph mobility *analytically*:

\[ n_{\Box}(c) = \frac{1}{2} N^2 \langle k \rangle^2 - \sum_j k_j^{in} k_j^{out} + \frac{1}{2} \text{Tr}(c^2) + \frac{1}{2} \text{Tr}(c^\dagger cc^\dagger c) + \text{Tr}(c^2 c^\dagger) - \sum_{ij} k_i^{in} c_{ij} k_j^{out} \]

*invariant*

\[ n_{\triangle}(c) = \frac{1}{3} \text{Tr}(c^3) - \text{Tr}(\hat{c}c^2) + \text{Tr}(\hat{c}^2 c) - \frac{1}{3} \text{Tr}(\hat{c}^3) \]

*state dependent*

*state dependent*

with: \((c^\dagger)_{ij} = c_{ji}, \hat{c}_{ij} = c_{ij}c_{ji}\)
Introduction
  Networks and graphs
  Tailored random graph ensembles

Counting tailored random graphs
  Entropy and complexity
  Nondirected graphs
  Directed graphs

Generating tailored random graphs
  Common algorithms and their problems
  MCMC processes for hard-constrained graphs

Degree-constrained MCMC graph dynamics
  Bookkeeping of moves
  Mobility of nondirected graphs
  Directed graphs

Tailoring loopy graph ensembles
  Motivation
  Spectrally constrained ensembles
  Solvable toy model
Motivation

is our ‘tailoring’ adequate?

e.g. do we recover phase transitions of Ising models on tailored random graphs?

- \( c^* = d\)-dim cubic lattice
  \( p(k) = \delta_{k,2^d} \)

- \( c^* = \) ‘small world’ lattice
  \( p(k \geq 2) = e^{-q} q^{k-2}/(k-2)! \)

\( \Omega_A: \) correct \( \langle k \rangle \)
\( \Omega_B: \) correct \( p(k) \)
\( \Omega_C: \) correct \( p(k) \) and \( W(k, k') \)

\( T_c(d) \)

\( T_c(1) = 0 \)
\( T_c(2) = 2/\log(1 + \sqrt{2}) \)
\( T_c(3) \approx 4.512 \)
\( T_c(4) \approx 6.687 \)

\( T_c(q) \)

\( T_c(0) = 0 \)
\( T_c(1) = 2/\log(1 + \sqrt{2}) \)
\( T_c(2) = 2/\log(3/4 + 1/4 \sqrt{17}) \)
\( T_c(3) = 2/\log(2/3 + 1/3 \sqrt{7}) \)
It is all about short loops ...

<table>
<thead>
<tr>
<th>degrees</th>
<th>4-loops</th>
<th>$d = 1$</th>
<th>$d = 2$</th>
<th>$d = 3$</th>
<th>$d = 4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>random, $\langle k \rangle = 2d$</td>
<td></td>
<td>1.820</td>
<td>3.915</td>
<td>5.944</td>
<td>7.958</td>
</tr>
<tr>
<td>random, $p(k) = \delta_{k,2d}$</td>
<td>✓</td>
<td>0</td>
<td>2.885</td>
<td>4.933</td>
<td>6.952</td>
</tr>
<tr>
<td>hypercubic Bethe</td>
<td>✓ ✓</td>
<td>0</td>
<td>2.771</td>
<td>4.839</td>
<td>6.879</td>
</tr>
<tr>
<td>true cubic lattice</td>
<td>✓ ✓</td>
<td>0</td>
<td>2.269</td>
<td>4.511</td>
<td>6.680</td>
</tr>
</tbody>
</table>

hypercubic Bethe lattice:  
‘tree of hypercubes’

- correct local degrees  
- geometric (non-random)  
- finite nr of short loops per site
maximum entropy random graphs with prescribed $p(k), W(k, k')$: locally tree-like ...

network | $n$ | $z$ | clustering coefficient $C$
--- | --- | --- | ---
Internet (autonomous systems)$^a$ | 6,374 | 3.8 | 0.24 | 0.00060
World-Wide Web (sites)$^b$ | 153,127 | 35.2 | 0.11 | 0.00023
power grid$^c$ | 4,941 | 2.7 | 0.080 | 0.00054
biology collaborations$^d$ | 1,520,251 | 15.5 | 0.081 | 0.000010
mathematics collaborations$^e$ | 253,339 | 3.9 | 0.15 | 0.000015
film actor collaborations$^f$ | 449,913 | 113.4 | 0.20 | 0.00025
company directors$^f$ | 7,673 | 14.4 | 0.59 | 0.0019
word co-occurrence$^g$ | 460,902 | 70.1 | 0.44 | 0.00015
neural network$^c$ | 282 | 14.0 | 0.28 | 0.049
metabolic network$^h$ | 315 | 28.3 | 0.59 | 0.090
food web$^i$ | 134 | 8.7 | 0.22 | 0.065

more realistic graph tailoring: constrain nr of short loops

problem: most analysis methods, e.g. replicas, GFA, cavity method, belief prop, etc require locally tree-like graphs (modulo loop corrections)

exceptions:
cubic lattices $d < 3$
spherical models
recent immune models
Spectrally constrained ensembles

- control closed paths of all lengths

\[ p(c) = \frac{1}{Z} \delta_{k,k(c)} e^{\sum_{\ell \geq 3} \alpha_{\ell} \sum_{i_1 \ldots i_\ell} c_{i_1} c_{i_2} \ldots c_{i_\ell} i_1} \]

generating function:

\[ \phi = \frac{1}{N} \log \sum_c \delta_{k,k(c)} e^{\sum_{\ell \geq 3} \alpha_{\ell} \operatorname{Tr}(c^\ell)} \]

\[ \langle m_\ell \rangle = \frac{1}{N} \langle \operatorname{Tr}(c^\ell) \rangle = \frac{\partial \phi}{\partial \alpha_{\ell}} \]

\[ S = \phi - \sum_{\ell \geq 3} v_\ell \langle m_\ell \rangle \]

- \( \operatorname{Tr}(c^\ell) = N \int d\mu \mu^\ell \varrho(\mu|c) \), so we control spectrum \( \varrho(\mu) \):

\[ p(c) = \frac{1}{Z} \delta_{k,k(c)} e^{N \int d\mu \hat{\varrho}(\mu) \varrho(\mu|c)} \]

\[ \hat{\varrho}(\mu) = \sum_{\ell \geq 3} \alpha_{\ell} \mu^\ell \]

generating function:

\[ \phi = \frac{1}{N} \log \sum_c \delta_{k,k(c)} e^{N \int d\mu \hat{\varrho}(\mu) \varrho(\mu|c)} \]

\[ \varrho(\mu) = \frac{\delta \phi}{\delta \hat{\varrho}(\mu)} \]

\[ S = \phi - \int d\mu \hat{\varrho}(\mu) \varrho(\mu) \]
Q1: How informative are spectra of finitely connected graphs?

Q2: How many non-isomorphic graphs are there with given degrees \((k_1, \ldots, k_N)\) and a given spectrum \(\varrho(\mu)\)?

Q3: How similar are processes running on non-isomorphic graphs with the same degrees \((k_1, \ldots, k_N)\) and the same spectrum \(\varrho(\mu)\)?

(spherical spins: free energies identical!)

how to compute

\[
\phi = \frac{1}{N} \log \sum_c \delta_{k,k(c)} e^{N \int d\mu \hat{\varrho}(\mu) \varrho(\mu | c)}
\]
Analytical route forward \[ p(\mathbf{c}) = \frac{1}{Z} \delta_{\mathbf{k}, \mathbf{k}(\mathbf{c})} e^{N \int d\mu \; \hat{\varphi}(\mu) \varrho(\mu|\mathbf{c})} \]

- Edwards-Jones:
\[
\varrho(\mu|\mathbf{c}) = \frac{2}{N\pi} \lim_{\varepsilon \downarrow 0} \text{Im} \frac{\partial}{\partial \mu} \log Z(\mu+i\varepsilon|\mathbf{c}), \quad Z(\mu|\mathbf{c}) = \int d\phi \; e^{-\frac{1}{2}i\phi \cdot [\mathbf{c}-\mu\mathbf{1}] \phi}
\]

- insert, integrate by parts, discretize $\mu$-integral:
\[
\phi = \frac{1}{N} \log \sum_{\mathbf{c}} \delta_{\mathbf{k}, \mathbf{k}(\mathbf{c})} e^{N \int d\mu \; \hat{\varphi}(\mu) \frac{2}{N\pi} \lim_{\varepsilon \downarrow 0} \text{Im} \frac{\partial}{\partial \mu} \log Z(\mu+i\varepsilon|\mathbf{c})}
\]
\[
= \lim_{\varepsilon, \Delta \downarrow 0} \frac{1}{N} \log \sum_{\mathbf{c}} \delta_{\mathbf{k}, \mathbf{k}(\mathbf{c})} \prod_{\mu} e^{-2\text{Im} \log Z(\mu+i\varepsilon|\mathbf{c})} \cdot \frac{\Delta}{\pi} \frac{d}{d\mu} \hat{\varphi}(\mu)
\]

- $e^{-2 \text{Im} \log z} = z^i \bar{z}^{-i}$

\[
\phi = \lim_{\varepsilon, \Delta \downarrow 0} \frac{1}{N} \log \sum_{\mathbf{c}} \delta_{\mathbf{k}, \mathbf{k}(\mathbf{c})} \prod_{\mu} \left[ Z(\mu+i\varepsilon|\mathbf{c})^i \frac{Z(\mu+i\varepsilon|\mathbf{c})^{-i}}{Z(\mu+i\varepsilon|\mathbf{c})} \right] \frac{\Delta}{\pi} \frac{d}{d\mu} \hat{\varphi}(\mu)
\]
\[ \phi = \lim_{\epsilon, \Delta \downarrow 0} \frac{1}{N} \log \sum_{\mathbf{c}} \delta_{k, k(c)} \prod_{\mu} \left[ Z(\mu + i\epsilon|\mathbf{c})^{n(\mu)} Z(\mu + i\epsilon|\mathbf{c})^{m(\mu)} \right], \]

\[ n(\mu) = \frac{i\Delta}{\pi} \frac{d}{d\mu} \hat{\varrho}(\mu), \]

\[ m(\mu) = -\frac{i\Delta}{\pi} \frac{d}{d\mu} \hat{\varrho}(\mu) \]

- replica method: factorization over entries \( \{c_{ij}\} \) (products of Gaussian integrals)

- steepest descent for \( N \to \infty \), continuation to imaginary dimensions, limits \( \epsilon \downarrow 0 \) and \( \Delta \downarrow 0 \)

- replica symmetry, bifurcation analysis, phase transitions and entropy

- elegant order parameter equations, interpretation in terms of ‘loopy’ message passing with a twist, treelike results (entropy, spectrum, ...) all recovered for \( \hat{\varrho}(\mu) \to 0 \)

but still wrong ... !?
Solvable toy model

simplest member of the family:

\[ k = (2, \ldots, 2), \quad \hat{\varnothing}(\mu) = \sum_{\ell=3}^{K} \alpha_{\ell} \mu^{\ell} : \quad p(c) = \frac{e^{\sum_{\ell=1}^{K} \alpha_{\ell} \text{Tr}(c^{\ell})}}{Z(\alpha)} \prod_{i=1}^{N} \delta_{\sum j c_{ij}, 2} \]

control nr of closed paths up to length \( K \) in 2-regular graphs ...

- all 2-regular graphs \( c \): collections of rings, combinatorics solvable:

\[
\lim_{N \to \infty} \phi_{N} = \lim_{N \to \infty} \text{extr}_{\omega} \left[ i\omega + \sum_{\ell=3}^{K} \frac{e^{(\alpha_{\ell} - i\omega)\ell}}{2\ell N} + \sum_{\ell=K+1}^{N} \frac{e^{-i\omega \ell}}{2\ell N} \right]
\]

densities \( m_{\ell} \) of length \( \ell \leq K \) closed paths

always vanish for \( N \to \infty \) ...

- Canonical parameter scaling: \( \alpha_{\ell} = \tilde{\alpha}_{\ell} + \ell^{-1} \log(N) \)

\[
\varphi(\tilde{\alpha}) = \lim_{N \to \infty} \text{extr}_{\omega} \left\{ i\omega + \sum_{\ell=3}^{K} \frac{e^{(\tilde{\alpha}_{\ell} - i\omega)\ell}}{2\ell} + \sum_{\ell=K+1}^{N} \frac{e^{-i\omega \ell}}{2\ell N} \right\}
\]
\[ p(c) = \frac{e^{\sum_{\ell=1}^{K} \left( \ell^{-1} \log N + \tilde{\alpha}_\ell \right) \text{Tr}(c^\ell)}}{Z(\alpha)} \prod_{i=1}^{N} \delta \sum_{j} c_{ij}, 2 \]

- **finite densities** \( m_\ell \) of closed paths

- **two phases**, critical manifold: \( 1 = \frac{1}{2} \sum_{\ell=3}^{K} e^{\ell \tilde{\alpha}_\ell} \)
  - disconnected (large \( \tilde{\alpha} \)): no extensively large rings, only small loops
  - connected (small \( \tilde{\alpha} \)): extensively large rings exist

\[ K = 3 \]

Simulations: \( N = 1000, 5000 \)

Solid line: theory
Lesson for spectrally constrained ensembles:

\[ p(c) = \frac{1}{Z} \delta_{k,k(c)} e^{N \int d\mu \, \tilde{\varrho}(\mu)c(\mu|c)}, \quad \tilde{\varrho}(\mu) \rightarrow \bar{\varrho}(\mu) \log N + \tilde{\varrho}(\mu) \]
Some references

Entropies of tailored graph ensembles

Graph dynamics and generation
Coolen, Annibale, Roberts,
*Generating random networks and graphs*
(Oxford University Press, 2017)

Loopy graph ensembles
Aguirre-Lopez, Barucca, Fekom, Coolen,
*arXiv:1705.03743, 2017*
https://nms.kcl.ac.uk/ton.coolen