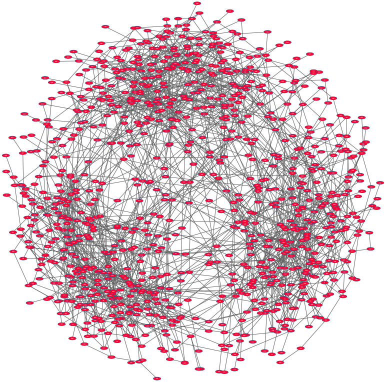
Tailoring of Random Networks and Graphs

ACC Coolen, King's College London

Eindhoven, Sept 21st 2017

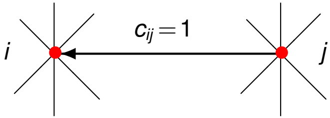


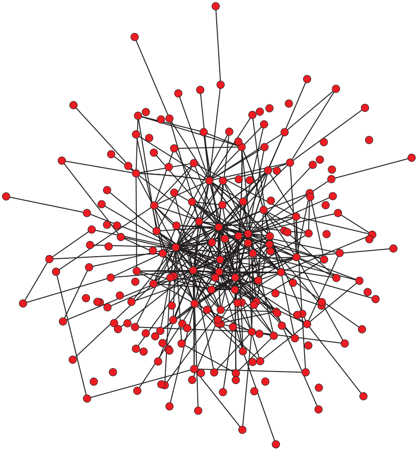
Networks and graphs

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: $\mathbf{c} = \{c_{ij}\}$





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 ...

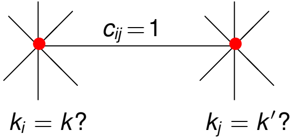
i.e. to have appropriate domain-specific statistical characteristics

Quantify topology of nondirected graphs

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

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

- joint degree statistics
  of connected nodes



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

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

$$\textit{assortativity/dissortativity}: \qquad C = \langle kk'\rangle_W - \langle k\rangle_W\langle k'\rangle_W$$

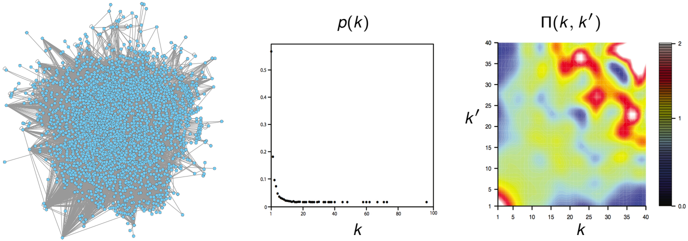
- marginals of W carry no info beyond degree statistics,

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

so focus on:

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

if $\exists(k,k')$ with $\Pi(k,k'|\mathbf{c}) \neq 1$:
structural information in degree correlations



human PIN
$N = 9306$
$\langle k\rangle = 7.53$

Quantify topology of directed graphs

links become *arrows*

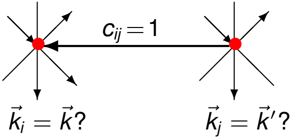
- degrees,
  degree sequences: $k_i^{\rm in}(\mathbf{c}) = \sum_j c_{ij}, \qquad \mathbf{k}^{\rm in}(\mathbf{c}) = (k_1^{\rm in}(\mathbf{c}), \ldots, k_N^{\rm in}(\mathbf{c}))$

  $$k_i^{\rm out}(\mathbf{c}) = \sum_j c_{ji}, \qquad \mathbf{k}^{\rm out}(\mathbf{c}) = (k_1^{\rm out}(\mathbf{c}), \ldots, k_N^{\rm out}(\mathbf{c}))$$

  degree distribution:

  $$k_i \;\to\; \vec{k}_i = (k_i^{\rm in}, k_i^{\rm out}) \qquad\qquad p(\vec{k}|\mathbf{c}) = \frac{1}{N}\sum_i \delta_{\vec{k},\vec{k}_i(\mathbf{c})}$$

- joint in-out degree statistics
  of connected nodes

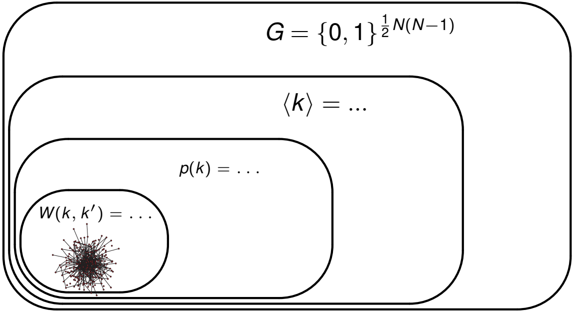


$$W(\vec{k}, \vec{k}'|\mathbf{c}) = \frac{1}{N\langle k\rangle}\sum_{ij} c_{ij}\,\delta_{\vec{k},\vec{k}_i(\mathbf{c})}\,\delta_{\vec{k}',\vec{k}_j(\mathbf{c})}$$

note:

$$W(\vec{k}, \vec{k}'|\mathbf{c}) \neq W(\vec{k}', \vec{k}|\mathbf{c})$$

Graph classification
via increasingly detailed
feature prescription



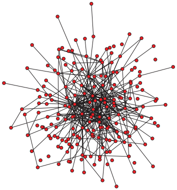
Tailoring random graphs

maximum entropy random graph ensembles,
$p(\mathbf{c})$ 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 $\mathbf{c}$? 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 $p(\mathbf{c})$ on G

- Tailoring via hard constraints

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

$$p(\mathbf{c}|\mathbf{\Omega}) = \frac{\delta_{\mathbf{\Omega}(\mathbf{c}),\mathbf{\Omega}}}{\mathcal{N}(\mathbf{\Omega})}, \qquad \mathcal{N}(\mathbf{\Omega}) = \sum_{\mathbf{c}} \delta_{\mathbf{\Omega}(\mathbf{c}),\mathbf{\Omega}} \quad (\#\ \textit{graphs in ensemble})$$

with $\mathbf{\Omega} = (\Omega_1, \ldots, \Omega_p)$

note:

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

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

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

- Tailoring via soft constraints

  impose *averages* for observables: $\Omega_\mu(\mathbf{c}) = \Omega_\mu$ for $\mu = 1 \ldots p$
  $p(\mathbf{c})$: maximum entropy, subject to constraints

$$p(\mathbf{c}|\boldsymbol{\Omega}) = Z^{-1}(\boldsymbol{\Omega})\, \mathrm{e}^{\sum_\mu \omega_\mu \Omega_\mu(\mathbf{c})}, \qquad Z(\boldsymbol{\Omega}) = \sum_{\mathbf{c}} \mathrm{e}^{\sum_\mu \omega_\mu \Omega_\mu(\mathbf{c})}$$

  parameters ω_μ:
  to be solved from

$$\forall \mu : \quad \sum_{\mathbf{c}} p(\mathbf{c}|\boldsymbol{\Omega}) \Omega_\mu(\mathbf{c}) = \Omega_\mu$$

  now *all* graphs **c** can emerge,
  but those with $\boldsymbol{\Omega}(\mathbf{c}) \approx \boldsymbol{\Omega}$ are most likely

  effective # graphs $\mathcal{N}(\boldsymbol{\Omega})$ defined via entropy:

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

Example

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

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

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

- calculate $\mathcal{N}(\langle k\rangle)$:

  use $\delta_{nm} = (2\pi)^{-1} \int_{-\pi}^{\pi} \mathrm{d}\omega\ \mathrm{e}^{\mathrm{i}(n-m)\omega}$

$$\begin{aligned}\mathcal{N}(\langle k\rangle) &= \int_{-\pi}^{\pi} \frac{\mathrm{d}\omega}{2\pi}\ \mathrm{e}^{\mathrm{i}\omega N\langle k\rangle} \sum_{\mathbf{c}} \mathrm{e}^{-\mathrm{i}\omega \sum_{ij} c_{ij}} = \binom{\frac{1}{2}N(N-1)}{\frac{1}{2}N\langle k\rangle} \\ &= \mathrm{e}^{\frac{1}{2}N\langle k\rangle \left[\log(N/\langle k\rangle)+1\right] + \mathcal{O}(\log N)}\end{aligned}$$

Example

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

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

$$p(\mathbf{c}|\langle k \rangle) = \frac{1}{Z(\omega)} \mathrm{e}^{\omega \sum_{ij} c_{ij}}, \qquad Z(\omega) = \sum_{\mathbf{c}} \mathrm{e}^{\omega \sum_{ij} c_{ij}}$$

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

- calculate $Z(\omega)$ and ω:

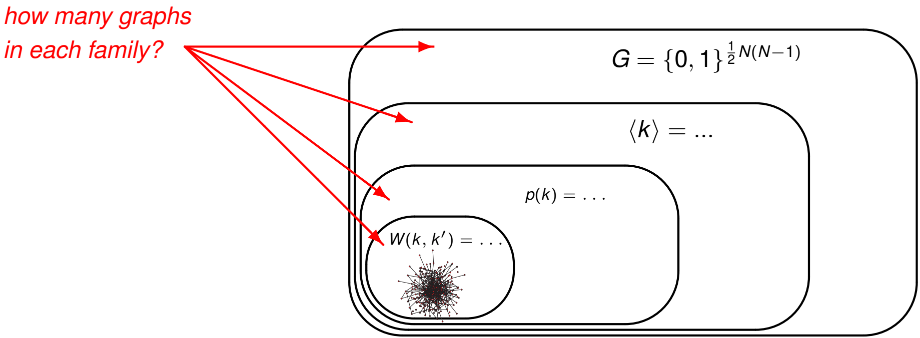
$$\langle k \rangle = (N-1) \frac{\mathrm{e}^{2\omega}}{\mathrm{e}^{2\omega}+1}$$

- rewrite probabilities:

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

Erdös-Rényi ensemble

Counting tailored random graphs



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 $p(\mathbf{c}|\mathbf{\Omega})$:
  ($\mathbf{\Omega}$: values of imposed observables)

$$\mathcal{N}(\mathbf{\Omega}) = \mathrm{e}^{N\langle k\rangle S(\mathbf{\Omega})}, \qquad S(\mathbf{\Omega}) = -\frac{1}{N\langle k\rangle}\sum_{\mathbf{c}\in G} p(\mathbf{c}|\mathbf{\Omega})\log p(\mathbf{c}|\mathbf{\Omega})$$

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

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

  $\exists$ many graphs with feature $\mathbf{\Omega}$: graphs with $\mathbf{\Omega}$ have *low complexity*
  $\exists$ few graphs with feature $\mathbf{\Omega}$: graphs with $\mathbf{\Omega}$ have *high complexity*

$$\mathcal{C}(\mathbf{\Omega}) = S(\emptyset) - S(\mathbf{\Omega})$$

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

$$p(\mathbf{c}|\emptyset) = 2^{-\frac{1}{2}N(N-1)}, \qquad S(\emptyset) = -\frac{1}{N\langle k\rangle}\log 2^{-\frac{1}{2}N(N-1)} = \frac{N-1}{2\langle k\rangle}\log 2$$

Nondirected graphs

$$p(\mathbf{c}) = \sum_{\mathbf{k}} \Big[\prod_i \mathrm{d}k_i \; p(k_i)\Big] \frac{\prod_i \delta_{k_i,k_i(\mathbf{c})}}{Z(\mathbf{k},W)} \prod_{i<j} \Big[\frac{\langle k\rangle}{N} \frac{W(k_i,k_j)}{p(k_i)p(k_j)} \delta_{c_{ij},1} + \Big(1 - \frac{\langle k\rangle}{N} \frac{W(k_i,k_j)}{p(k_i)p(k_j)}\Big) \delta_{c_{ij},0}\Big]$$

$$S = \underbrace{\frac{1}{2}[1+\log(\frac{N}{\langle k\rangle})]}_{Erdos-Renyi\ \ entropy} - \Big\{ \underbrace{\frac{1}{\langle k\rangle} \sum_k p(k) \log[\frac{p(k)}{\tilde{p}(k)}]}_{degree\ \ complexity} + \underbrace{\frac{1}{2} \sum_{k,k'} W(k,k') \log\Big[\frac{W(k,k')}{W(k)W(k')}\Big]}_{wiring\ \ complexity} \Big\}$$

$$+ \epsilon_N$$

$\lim_{N\to\infty} \epsilon_N = 0$

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

degree distr of Erdös-Renyi graphs

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

Directed graphs

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

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

$$S = \underbrace{1 + \log(\frac{N}{\langle k\rangle})}_{directed\ \ ER\ \ entropy} - \Big\{ \underbrace{\frac{1}{\langle k\rangle} \sum_{\vec{k}} p(\vec{k}) \log[\frac{p(\vec{k})}{\tilde{p}(k^{\rm in})\tilde{p}(k^{\rm out})}]}_{degree\ \ complexity} + \underbrace{\sum_{\vec{k},\vec{k}'} W(\vec{k}, \vec{k}') \log\Big[\frac{W(\vec{k}, \vec{k}')}{W(\vec{k})W(\vec{k}')}\Big]}_{wiring\ \ complexity} \Big\}$$

$$+ \epsilon_N$$

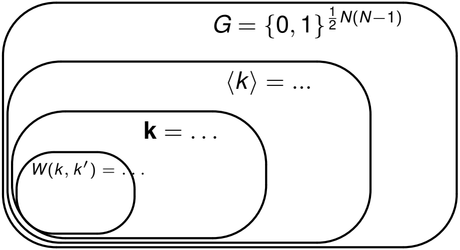
$\lim_{N\to\infty} \epsilon_N = 0$

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

$\tilde{p}(k^{\rm in})\tilde{p}(k^{\rm out})$: degree distr of *directed* Erdös-Renyi graphs

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

Generating tailored random graphs



G: all nondirected N-node graphs
$G[\mathbf{k}] \subset G$: all nondirected N-node graphs with degrees $\mathbf{k}$

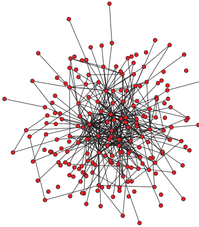
typical questions:
how to generate numerically

- random $\mathbf{c} \in G$, with *specified* $p(\mathbf{c})$
- random $\mathbf{c} \in G[\mathbf{k}]$, with *uniform* $p(\mathbf{c})$
- random $\mathbf{c} \in G[\mathbf{k}]$, with *specified* $p(\mathbf{c})$

similar for directed graphs ...

Common algorithms and their problems

soft constraints only:
standard Glauber/Gibbs/MCMC dynamics



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

strategy: start from any graph $\mathbf{c}$
propose random moves $c_{ij} \to 1-c_{ij}$ (giving $\mathbf{c} \to F_{ij}\mathbf{c}$),

define acceptance probabilities $A(F_{ij}\mathbf{c}|\mathbf{c})$
via detailed balance condition

$$A(F_{ij}\mathbf{c}|\mathbf{c})p(\mathbf{c}) = A(\mathbf{c}|F_{ij}\mathbf{c})p(F_{ij}\mathbf{c}) \quad \to \quad A(\mathbf{c}'|\mathbf{c}) = \Big[1 + p(\mathbf{c})/p(\mathbf{c}')\Big]^{-1}$$

stochastic process is ergodic, and converges to $p(\mathbf{c})$

practicalities:
equilibration can take a *very long* time,
so monitor Hamming distances

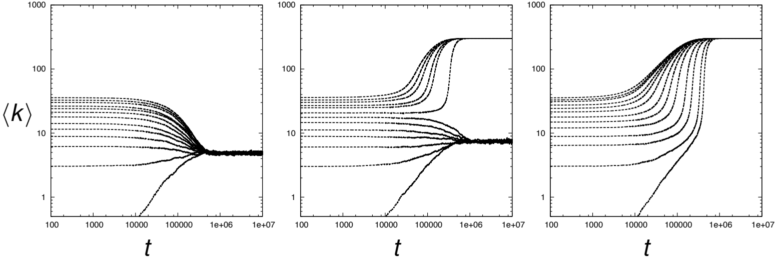
similar for directed graphs ...

The problem of phase transitions

example:

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

$\beta = 0.01$ $\beta = 0.02$ $\beta = 0.03$



- 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 $\mathbf{c} \in \{0,1\}^{\frac{1}{2}N(N-1)}$
with specified degree sequence $\mathbf{k} = (k_1, \ldots, k_N)$

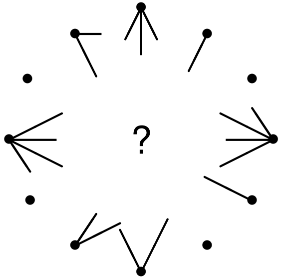
strategy: stochastic growth dynamics,
starting from graph with no links

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

repeat:

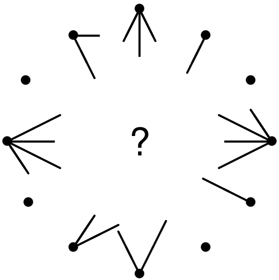
- pick at random two nodes (i,j)
- if $\sum_\ell c_{i\ell} < k_i$ and $\sum_\ell c_{j\ell} < k_j$:
  connect i and j
  $c_{ij} = 0 \;\rightarrow\; c_{ij} = 1$

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 $\mathbf{c} \in G[\mathbf{k}]$,
  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

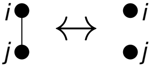
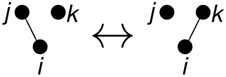
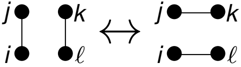
- sampling bias:

  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 $p(\mathbf{c})$ is flat ...

MCMC with hard constraints

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

MOVE SET	INVARIANTS	ACTION
Link flips $\{F_{ij}\}$	none	
Hinge flips $\{F_{ijk}\}$	average degree $\bar{k}(\mathbf{c}) = \frac{1}{N}\sum_{rs} c_{rs}$	
Edge swaps $\{F_{ijk\ell}\}$	all individual degrees $k_i(\mathbf{c}) = \sum_j c_{ij},\ \ i = 1 \dots N$	

Edge switching algorithm
(Seidel, 1976)

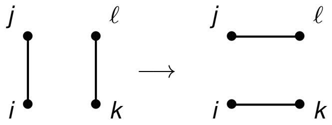
objective: generate random nondirected graph $\mathbf{c} \in \{0,1\}^{\frac{1}{2}N(N-1)}$
with specified degree sequence $\mathbf{k} = (k_1, \ldots, k_N)$

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:

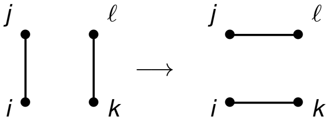
- pick at random four nodes (i, j, k, ℓ)
  that are *pairwise connected*
- carry out an 'edge swap'
  (or 'Seidel switch), see diagram
  (preserves all degrees!)



terminate if stochastic process has equilibrated

Edge switching algorithm

limitations and problems ...



- major limitation:

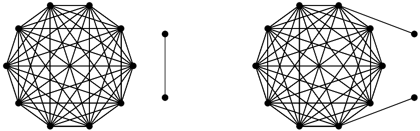
  aims to generate random $\mathbf{c} \in G[\mathbf{k}]$,
  but cannot control graph probabilities ...

- inconvenience: need for a 'seed graph'
  with the correct degrees $\mathbf{k} = (k_1, \ldots, k_N)$

- sampling bias:

  edge swaps are ergodic on $G[\mathbf{k}]$ (Taylor, 1981),
  but sampling is *not uniform*!

many possible moves *few moves ...*



nr of possible moves
depends on state **c**!

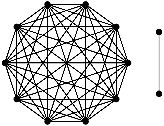
result:
stationary state of Markov chain
favours high-mobility graphs

dangerous for scale-free graphs ...

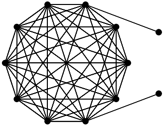
target:
uniform measure $p(\mathbf{c})$
on $G[\mathbf{k}]$

$n(\mathbf{c})$: nr of possible moves

1 *graph*
$n(\mathbf{c}) = (N-2)(N-3)$



$(N-2)(N-3)$ *graphs*
$n(\mathbf{c}) = 2(N-3)$

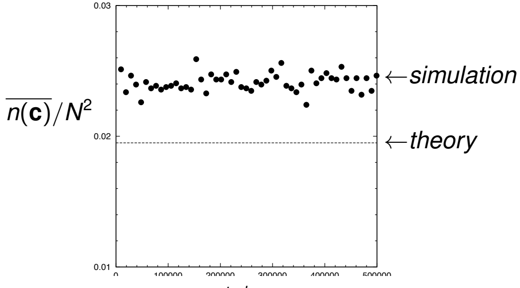


for flat measure:

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

$N = 100$:
$\overline{n(\mathbf{c})}/N^2 \approx 0.0195$

'accept all'
edge swapping:



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:
  $G[\star] \subseteq G$: all $\mathbf{c} \in G$ that satisfy constraints $\star$

- stochastic graph dynamics as a Markov chain,
  transition probabilities $W(\mathbf{c}|\mathbf{c}')$ for move $\mathbf{c}' \to \mathbf{c}$

$$\forall \mathbf{c} \in G[\star]: \qquad p_{t+1}(\mathbf{c}) = \sum_{\mathbf{c}' \in G[\star]} W(\mathbf{c}|\mathbf{c}')p_t(\mathbf{c}')$$

- allowed moves (exclude identity):

  Φ: set of allowed moves $F: G_F[\star] \to G[\star]$
  $G_F[\star]$: those $\mathbf{c} \in G[\star]$ on which F can act

  all moves are auto-invertible: $(\forall F \in \Phi): \; F^2 = \mathbb{I}$
  Φ is ergodic on $G[\star]$

objective

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

- standard form:

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

$$\begin{array}{rl} q(F|\mathbf{c}): & \textit{move proposal probability} \\ A(\mathbf{c}|\mathbf{c}'): & \textit{move acceptance probability} \end{array}$$

detailed balance:

$$(\forall F\in\Phi)(\forall \mathbf{c}\in G[\star]): \qquad q(F|\mathbf{c})A(F\mathbf{c}|\mathbf{c})\mathrm{e}^{-H(\mathbf{c})} = q(F|F\mathbf{c})A(\mathbf{c}|F\mathbf{c})\mathrm{e}^{-H(F\mathbf{c})}$$

- move proposal probability:

$$q(F|\mathbf{c}) = \begin{cases} 0 & \textit{if } F \textit{ cannot act on } \mathbf{c} \\ 1/n(\mathbf{c}) & \textit{if } F \textit{ can act on } \mathbf{c} \end{cases}$$

graph mobility $n(\mathbf{c})$:

$$n(\mathbf{c}) = \sum_{F\in\Phi} I_F(\mathbf{c}), \qquad I_F(\mathbf{c}) = \begin{cases} 1 & \textit{if } \mathbf{c}\in G_F[\star] \\ 0 & \textit{if } \mathbf{c}\notin G_F[\star] \end{cases}$$

canonical Markov chain

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

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

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

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

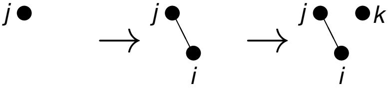
corresponds to
$H(\mathbf{c}) = -\log n(\mathbf{c})$,
so would give

$$\textit{sampling bias}: \quad p(\mathbf{c}) = \frac{n(\mathbf{c})}{\sum_{\mathbf{c}' \in G[\star]} n(\mathbf{c}')}$$

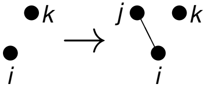
picking moves randomly ...

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

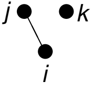
PROTOCOL 1:
(i) pick a site j with $k_j(\mathbf{A}) > 0$
(ii) pick a site $i \in \partial_j(\mathbf{A})$
(iii) pick a site $k \notin \partial_i(\mathbf{A}) \cup \{i\}$



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



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

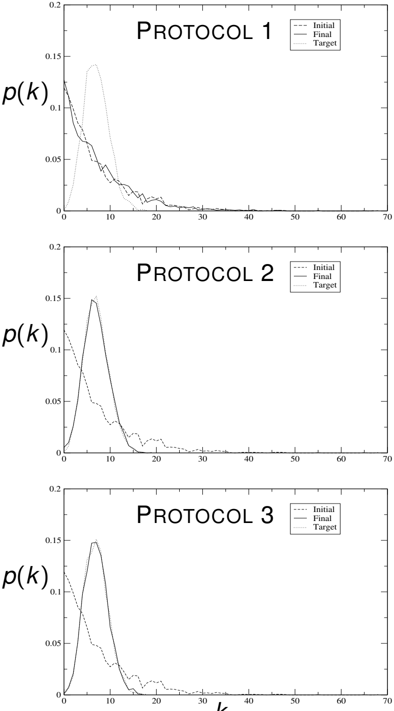


$N = 3000$, $\langle k \rangle = 7$

dashed: start graph

dotted: $p(k)$ of target $p(\mathbf{A})$

solid: MCMC result

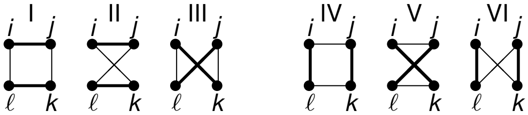


Bookkeeping of moves

- constraints: imposed degrees $\mathbf{k}$

  ergodic set Φ of admissible moves:
  edge swaps $F: G_F[\mathbf{k}] \to G[\mathbf{k}]$

  $\{(i,j,k,\ell) \in \{1,\ldots,N\}^4 |\, i<j<k<\ell\}$, 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 $\alpha \in \{1,2,3\}$

  $I_{ijk\ell;\alpha}(\mathbf{c}) = 1$:

$$\begin{aligned} F_{ijk\ell;\alpha}(\mathbf{c})_{qr} &= 1 - c_{qr} \quad \text{for } (q,r) \in \mathcal{S}_{ijk\ell;\alpha} \\ F_{ijk\ell;\alpha}(\mathbf{c})_{qr} &= c_{qr} \quad \text{for } (q,r) \notin \mathcal{S}_{ijk\ell;\alpha} \end{aligned}$$

$$\mathcal{S}_{ijk\ell;1} = \{(i,j),(k,\ell),(i,\ell),(j,k)\}, \quad \mathcal{S}_{ijk\ell;2} = \{(i,j),(k,\ell),(i,k),(j,\ell)\}$$
$$\mathcal{S}_{ijk\ell;3} = \{(i,k),(j,\ell),(i,\ell),(j,k)\}$$

Mobility of nondirected graphs

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

$n(\mathbf{c}) = \sum_{i<j<k<\ell}^{N} \sum_{\alpha=1}^{3} I_{ijk\ell;\alpha}(\mathbf{c})$

$$\begin{aligned}
I_{ijk\ell;1}(\mathbf{c}) &= c_{ij}c_{k\ell}(1-c_{i\ell})(1-c_{jk}) + (1-c_{ij})(1-c_{k\ell})c_{i\ell}c_{jk} \\
I_{ijk\ell;2}(\mathbf{c}) &= c_{ij}c_{k\ell}(1-c_{ik})(1-c_{j\ell}) + (1-c_{ij})(1-c_{k\ell})c_{ik}c_{j\ell} \\
I_{ijk\ell;3}(\mathbf{c}) &= c_{ik}c_{j\ell}(1-c_{i\ell})(1-c_{jk}) + (1-c_{ik})(1-c_{j\ell})c_{i\ell}c_{jk}
\end{aligned}$$

work out combinatorics:

$$n(\mathbf{c}) = \underbrace{\frac{1}{4}N^2\langle k\rangle^2 + \frac{1}{4}N\langle k\rangle - \frac{1}{2}N\langle k^2\rangle}_{invariant} + \underbrace{\frac{1}{4}\mathrm{Tr}(\mathbf{c}^4) + \frac{1}{2}\mathrm{Tr}(\mathbf{c}^3) - \frac{1}{2}\sum_{ij} k_i c_{ij} k_j}_{state\ dependent}$$

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

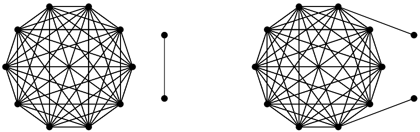
Example:

target =
uniform measure on $G[\mathbf{k}]$

$N = 100$

naive versus correct
acceptance probabilities

many possible moves *few moves*



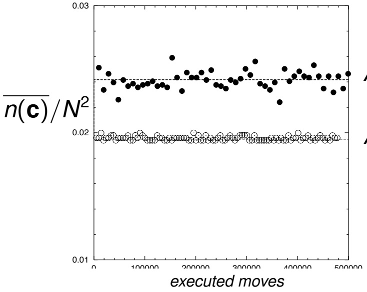
predictions:

$p(\mathbf{c}) = \textit{constant}$:

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

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

$\overline{n(\mathbf{c})}/N^2 \approx 0.0242$

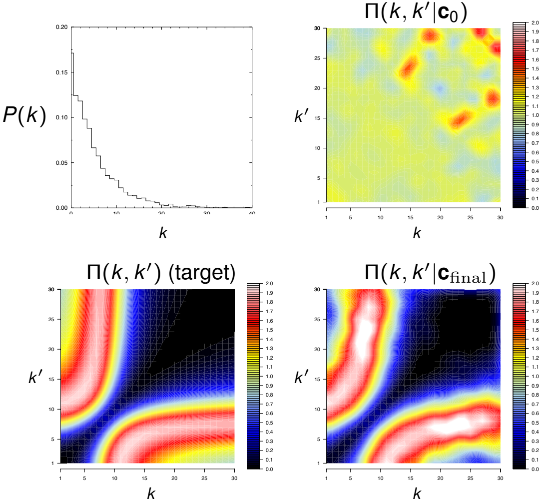


$A(\mathbf{c}|\mathbf{c}') = 1$

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

Example

target =
degree-correlated
measure on $G[\mathbf{k}]$



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

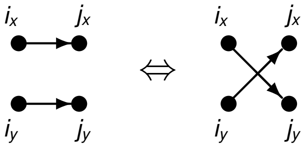
$$\Pi(k,k') = \frac{(k-k')^2}{[\beta_1 - \beta_2 k + \beta_3 k^2][\beta_1 - \beta_2 k' + \beta_3 k'^2]}$$

Directed graphs

bookkeeping of elementary moves

- constraints: imposed in-out degrees, so graph set is $G[\mathbf{k}^{\mathrm{in}}, \mathbf{k}^{\mathrm{out}}]$

  set Φ of admissible moves:
  directed edge swaps $\quad F : G_F[\mathbf{k}^{\mathrm{in}}, \mathbf{k}^{\mathrm{out}}] \to G[\mathbf{k}^{\mathrm{in}}, \mathbf{k}^{\mathrm{out}}]$

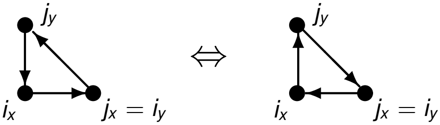


for *nondirected* graphs:
edge swaps are *ergodic* set of moves
(Taylor, 1981 – proof based on Lyapunov function)

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*:

$$n_\square(\mathbf{c}) = \underbrace{\frac{1}{2}N^2\langle k\rangle^2 - \sum_j k_j^{\rm in}k_j^{\rm out}}_{invariant} + \underbrace{\frac{1}{2}\mathrm{Tr}(\mathbf{c}^2) + \frac{1}{2}\mathrm{Tr}(\mathbf{c}^\dagger\mathbf{c}\mathbf{c}^\dagger\mathbf{c}) + \mathrm{Tr}(\mathbf{c}^2\mathbf{c}^\dagger) - \sum_{ij} k_i^{\rm in}c_{ij}k_j^{\rm out}}_{state\ dependent}$$

$$n_\triangle(\mathbf{c}) = \underbrace{\frac{1}{3}\mathrm{Tr}(\mathbf{c}^3) - \mathrm{Tr}(\hat{\mathbf{c}}\mathbf{c}^2) + \mathrm{Tr}(\hat{\mathbf{c}}^2\mathbf{c}) - \frac{1}{3}\mathrm{Tr}(\hat{\mathbf{c}}^3)}_{state\ dependent}$$

with: $(\mathbf{c}^\dagger)_{ij} = c_{ji},\ \hat{\mathbf{c}}_{ij} = c_{ij}c_{ji}$

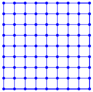
Motivation

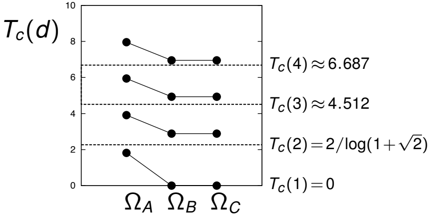
is our 'tailoring' adequate?

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

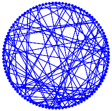
Ω_A: correct $\langle k \rangle$
Ω_B: correct $p(k)$
Ω_C: correct $p(k)$ and $W(k,k')$

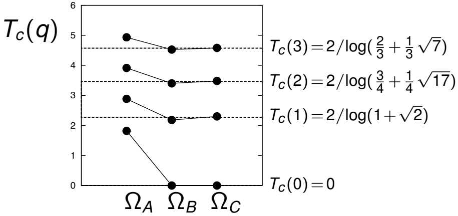
- $\mathbf{c}^\star$ = d-dim cubic lattice
  $p(k) = \delta_{k,2^d}$





- $\mathbf{c}^\star$ = 'small world' lattice
  $p(k \geq 2) = \mathrm{e}^{-q} q^{k-2}/(k-2)!$





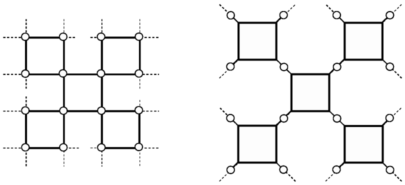
It is all about short loops ...

critical temperatures $T_c(d)$

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

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 measured	random graph
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*

$$p(\mathbf{c}) = \frac{1}{Z}\delta_{\mathbf{k},\mathbf{k}(\mathbf{c})}\ \mathrm{e}^{\sum_{\ell\geq 3}\alpha_\ell \sum_{i_1\ldots i_\ell} c_{i_1 i_2} c_{i_2 i_3}\ldots c_{i_\ell i_1}}$$

generating function:

$$\phi = \frac{1}{N}\log\sum_{\mathbf{c}}\delta_{\mathbf{k},\mathbf{k}(\mathbf{c})}\ \mathrm{e}^{\sum_{\ell\geq 3}\alpha_\ell \mathrm{Tr}(\mathbf{c}^\ell)}$$

$$\langle m_\ell\rangle = \tfrac{1}{N}\langle \mathrm{Tr}(\mathbf{c}^\ell)\rangle = \partial\phi/\partial\alpha_\ell$$

$$S = \phi - \textstyle\sum_{\ell\geq 3} v_\ell \langle m_\ell\rangle$$

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

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

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

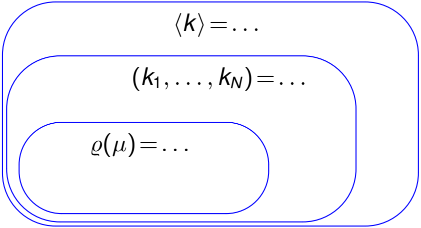
generating function:

$$\phi = \frac{1}{N}\log\sum_{\mathbf{c}}\delta_{\mathbf{k},\mathbf{k}(\mathbf{c})}\ \mathrm{e}^{N\int\mathrm{d}\mu\ \hat{\varrho}(\mu)\varrho(\mu|\mathbf{c})}$$

$$\varrho(\mu) = \delta\phi/\delta\hat{\varrho}(\mu)$$

$$S = \phi - \int\mathrm{d}\mu\ \hat{\varrho}(\mu)\varrho(\mu)$$

Some relevant questions



- 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

$$\phi = \frac{1}{N} \log \sum_{\mathbf{c}} \delta_{\mathbf{k},\mathbf{k}(\mathbf{c})} \, \mathrm{e}^{N \int \mathrm{d}\mu \ \hat{\varrho}(\mu) \varrho(\mu|\mathbf{c})}$$

Analytical route forward

$$p(\mathbf{c}) = \frac{1}{Z}\,\delta_{\mathbf{k},\mathbf{k}(\mathbf{c})}\;\mathrm{e}^{N\int \mathrm{d}\mu\;\hat{\varrho}(\mu)\varrho(\mu|\mathbf{c})}$$

- Edwards-Jones:

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

- insert, integrate by parts, discretize μ-integral:

$$\begin{aligned}\phi &= \frac{1}{N}\log\sum_{\mathbf{c}}\delta_{\mathbf{k},\mathbf{k}(\mathbf{c})}\;\mathrm{e}^{N\int \mathrm{d}\mu\;\hat{\varrho}(\mu)\frac{2}{N\pi}\lim_{\varepsilon\downarrow 0}\mathrm{Im}\frac{\partial}{\partial\mu}\log Z(\mu+\mathrm{i}\varepsilon|\mathbf{c})}\\ &= \lim_{\varepsilon,\Delta\downarrow 0}\frac{1}{N}\log\sum_{\mathbf{c}}\delta_{\mathbf{k},\mathbf{k}(\mathbf{c})}\prod_{\mu}\mathrm{e}^{-2\mathrm{Im}\log Z(\mu+\mathrm{i}\varepsilon|\mathbf{c}).\;\frac{\Delta}{\pi}\frac{\mathrm{d}}{\mathrm{d}\mu}\hat{\varrho}(\mu)}\end{aligned}$$

- $\mathrm{e}^{-2\;\mathrm{Im}\log z} = z^{\mathrm{i}}.\overline{z}^{\;-\mathrm{i}}$

$$\phi = \lim_{\varepsilon,\Delta\downarrow 0}\frac{1}{N}\log\sum_{\mathbf{c}}\delta_{\mathbf{k},\mathbf{k}(\mathbf{c})}\prod_{\mu}\Big[Z(\mu+\mathrm{i}\varepsilon|\mathbf{c})^{\mathrm{i}}\;\overline{Z(\mu+\mathrm{i}\varepsilon|\mathbf{c})}^{-\mathrm{i}}\Big]^{\frac{\Delta}{\pi}\frac{\mathrm{d}}{\mathrm{d}\mu}\hat{\varrho}(\mu)}$$

$$\phi = \lim_{\varepsilon,\Delta\downarrow 0} \frac{1}{N} \log \sum_{\mathbf{c}} \delta_{\mathbf{k},\mathbf{k}(\mathbf{c})} \prod_{\mu} \Big[Z(\mu+\mathrm{i}\varepsilon|\mathbf{c})^{n(\mu)}\, \overline{Z(\mu+\mathrm{i}\varepsilon|\mathbf{c})}^{m(\mu)} \Big],$$

$$n(\mu) = \frac{\mathrm{i}\Delta}{\pi} \frac{\mathrm{d}}{\mathrm{d}\mu} \hat{\varrho}(\mu)$$
$$m(\mu) = -\frac{\mathrm{i}\Delta}{\pi} \frac{\mathrm{d}}{\mathrm{d}\mu} \hat{\varrho}(\mu)$$

- 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 familiy:

$$\mathbf{k} = (2,\ldots,2), \quad \hat{\varrho}(\mu) = \sum_{\ell=3}^{K} \alpha_\ell \mu^\ell: \qquad p(\mathbf{c}) = \frac{\mathrm{e}^{\sum_{\ell=1}^{K} \alpha_\ell \mathrm{Tr}(\mathbf{c}^\ell)}}{Z(\boldsymbol{\alpha})} \prod_{i=1}^{N} \delta_{\sum_j c_{ij},2}$$

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

- all 2-regular graphs $\mathbf{c}$: collections of *rings*, combinatorics solvable:

$$\lim_{N\to\infty} \phi_N = \lim_{N\to\infty} \mathrm{extr}_\omega \Big[\mathrm{i}\omega + \sum_{\ell=3}^{K} \frac{\mathrm{e}^{(\alpha_\ell - \mathrm{i}\omega)\ell}}{2\ell N} + \sum_{\ell=K+1}^{N} \frac{\mathrm{e}^{-\mathrm{i}\omega\ell}}{2\ell N} \Big]$$

densities m_ℓ of length $\ell \le K$ closed paths
always vanish for $N \to \infty$ ...

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

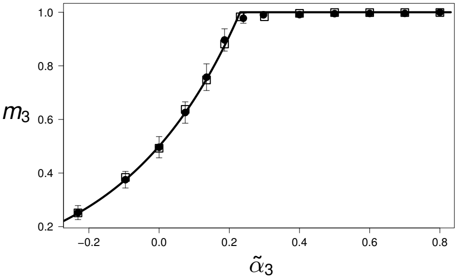
$$\varphi(\tilde{\boldsymbol{\alpha}}) = \lim_{N\to\infty} \mathrm{extr}_\omega \Big\{ \mathrm{i}\omega + \sum_{\ell=3}^{K} \frac{\mathrm{e}^{(\tilde{\alpha}_\ell - \mathrm{i}\omega)\ell}}{2\ell} + \sum_{\ell=K+1}^{N} \frac{\mathrm{e}^{-\mathrm{i}\omega\ell}}{2\ell N} \Big\}$$

$$p(\mathbf{c}) = \frac{\mathrm{e}^{\sum_{\ell=1}^{K}\left(\ell^{-1}\log N + \tilde{\alpha}_\ell\right)\mathrm{Tr}(\mathbf{c}^\ell)}}{Z(\boldsymbol{\alpha})} \prod_{i=1}^{N} \delta_{\sum_j c_{ij},2}$$

- *finite densities* m_ℓ of closed paths
- two *phases*, critical manifold: $1 = \frac{1}{2}\sum_{\ell=3}^{K} \mathrm{e}^{\ell\tilde{\alpha}_\ell}$

disconnected (*large* $\tilde{\alpha}$): *no extensively large rings*, *only small loops*

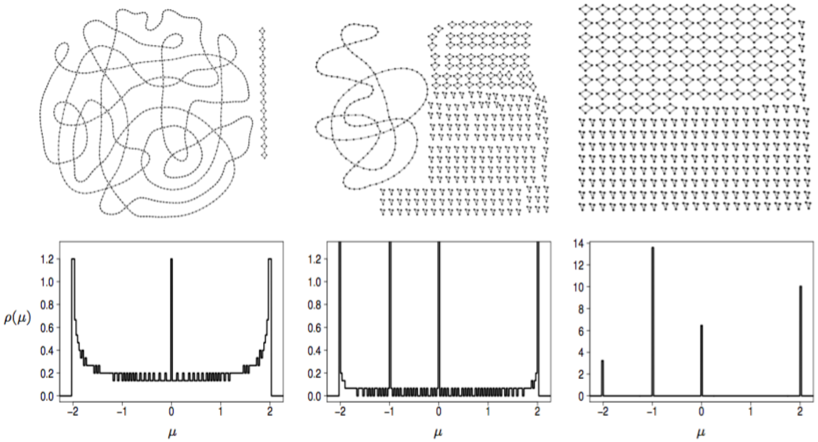
connected (*small* $\tilde{\alpha}$): *extensively large rings exist*



$K = 3$

simulations: $N = 1000$, 5000

solid line: theory



lesson for spectrally constrained ensembles:

$$p(\mathbf{c}) = \frac{1}{Z}\,\delta_{\mathbf{k},\mathbf{k}(\mathbf{c})}\,\mathrm{e}^{N\int \mathrm{d}\mu\;\hat{\varrho}(\mu)\varrho(\mu|\mathbf{c})}, \qquad \hat{\varrho}(\mu) \to \bar{\varrho}(\mu)\log N + \tilde{\varrho}(\mu)$$

Some references

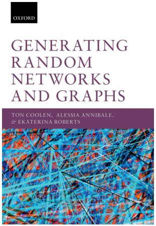
Entropies of tailored graph ensembles

Bianconi, Coolen, Perez-Vicente, *Phys Rev E 78, 2008*
Annibale, Coolen, Fernandes, Fraternali, Kleinjung, *J Phys A 42, 2009*
Roberts, Schlitt, Coolen, *J Phys A 44, 2011*
Roberts, Coolen, *J Phys A 47, 2014*

Graph dynamics and generation

Taylor, *Combinatorial Mathematics VIII, Springer Lect Notes Math 884, 1981*
Rao, Jana, Bandyopadhya, *Indian J of Statistics 58, 1996*
Coolen, De Martino, Annibale, *J Stat Phys 136, 2009*
Roberts, Coolen, *Phys Rev E 85, 2012*

Coolen, Annibale, Roberts,
Generating random networks and graphs
(Oxford University Press, 2017)



Loopy graph ensembles

Coolen, *J Phys Conference Series 699, 2016*
Aguirre-Lopez, Barucca, Fekom, Coolen,
arXiv:1705.03743, 2017

https://nms.kcl.ac.uk/ton.coolen