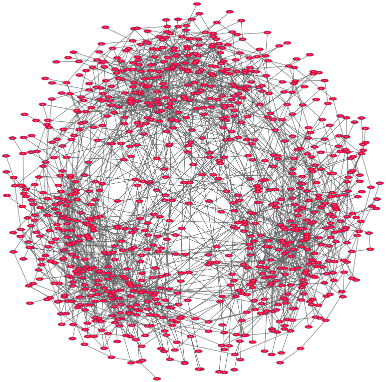
Statistical physics of tailored random graphs: entropies, processes, and generation

II. Tailored sparse random graphs

ACC Coolen, King's College London



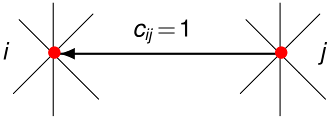
Background - tailored random graphs

networks/graphs:

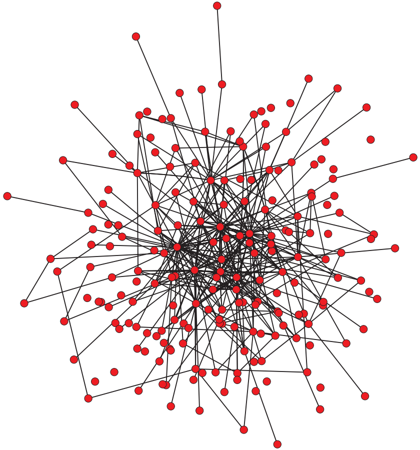
number of nodes: N
nodes (vertices): $i, j \in \{1, \ldots, N\}$

links (edges): $c_{ij} \in \{0, 1\}$
no self-links: $c_{ii} = 0$ for all i

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



nondirected graph: $\forall (i, j): \; c_{ij} = c_{ji}$
directed graph: $\exists (i, j): \; c_{ij} \neq c_{ji}$



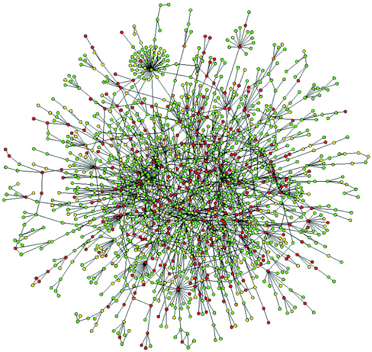
if we model real-world systems by graphs
we want these graphs to be realistic ...

Networks in cell biology

- **protein interaction networks**:

  nodes: proteins $i, j = 1 \ldots N$
  links: $c_{ij} = c_{ji} = 1$ if i can bind to j
  $c_{ij} = c_{ji} = 0$ otherwise

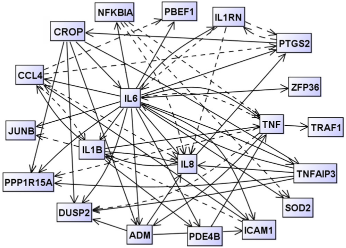
  nondirected graphs,
  $N \sim 10^4$, links/node ~ 7



- **gene regulation networks**:

  nodes: genes $i, j = 1 \ldots N$
  links: $c_{ij} = 1$ if j is transcription factor of i
  $c_{ij} = 0$ otherwise

  directed graphs,
  $N \sim 10^4$, links/node ~ 5

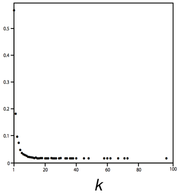


Quantify topology of nondirected graphs

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

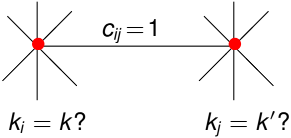
- degree distribution:

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



- joint degree statistics
  of connected nodes

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



normalisation:

$$\sum_{k,k'\geq 0} W(k,k'|\mathbf{c}) = \frac{1}{N\langle k\rangle}\sum_{ij} c_{ij} = \frac{1}{N\langle k\rangle}\sum_i k_i(\mathbf{c}) = 1$$

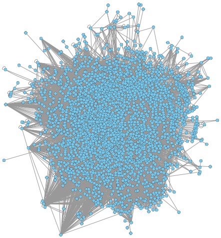
- relation between p and W:

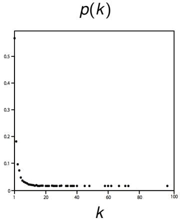
$$
\begin{aligned}
W(k|\mathbf{c}) &= \sum_{k'} W(k,k'|\mathbf{c}) = \frac{1}{N\langle k\rangle}\sum_{ij} c_{ij}\delta_{k,k_i(\mathbf{c})} \\
&= \frac{1}{N\langle k\rangle}\sum_{i} k_i(\mathbf{c})\delta_{k,k_i(\mathbf{c})} = \frac{k}{N\langle k\rangle}\sum_{i}\delta_{k,k_i(\mathbf{c})} = p(k|\mathbf{c})k/\langle k\rangle
\end{aligned}
$$

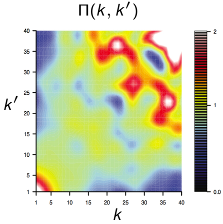
- hence maginals of W carry no info beyond degree statistics
  so focus on:

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

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







H sapiens PIN
$N = 9306$
$\langle k\rangle = 7.53$

Quantify topology of directed graphs

links now become *arrows*

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

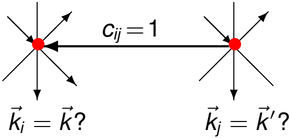
  $k_i^{\rm out}(\mathbf{c})=\sum_j c_{ji}$, $\quad \mathbf{k}^{\rm out}(\mathbf{c})=(k_1^{\rm out}(\mathbf{c}),\ldots,k_N^{\rm out}(\mathbf{c}))$

- 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:

$$W(\vec{k},\vec{k}'|\mathbf{c})-W(\vec{k}',\vec{k}|\mathbf{c})=\frac{1}{N\langle k\rangle}\sum_{ij}(c_{ij}-c_{ji})\;\delta_{\vec{k},\vec{k}_i(\mathbf{c})}\delta_{\vec{k}',\vec{k}_j(\mathbf{c})}\;\neq 0$$

- relation between p and W:

$$
\begin{aligned}
W_1(\vec{k}|\mathbf{c}) &= \sum_{\vec{k}'}\frac{1}{N\langle k\rangle}\sum_{ij}c_{ij}\delta_{\vec{k},\vec{k}_i(\mathbf{c})}\delta_{\vec{k}',\vec{k}_j(\mathbf{c})} = \frac{1}{N\langle k\rangle}\sum_{ij}c_{ij}\delta_{\vec{k},\vec{k}_i(\mathbf{c})} \\
&= \frac{1}{N\langle k\rangle}\sum_i k_i^{\rm in}(\mathbf{c})\delta_{\vec{k},\vec{k}_i(\mathbf{c})} = \frac{k^{\rm in}}{N\langle k\rangle}\sum_i \delta_{\vec{k},\vec{k}_i(\mathbf{c})} = p(\vec{k}|\mathbf{c})k^{\rm in}/\langle k\rangle
\end{aligned}
$$

$$
\begin{aligned}
W_2(\vec{k}'|\mathbf{c}) &= \sum_{\vec{k}}\frac{1}{N\langle k\rangle}\sum_{ij}c_{ij}\delta_{\vec{k},\vec{k}_i(\mathbf{c})}\delta_{\vec{k}',\vec{k}_j(\mathbf{c})} = \frac{1}{N\langle k\rangle}\sum_{ij}c_{ij}\delta_{\vec{k}',\vec{k}_j(\mathbf{c})} \\
&= \frac{1}{N\langle k\rangle}\sum_j k_j^{\rm out}(\mathbf{c})\delta_{\vec{k}',\vec{k}_j(\mathbf{c})} = \frac{k^{\rm out\prime}(\mathbf{c})}{N\langle k\rangle}\sum_j \delta_{\vec{k}',\vec{k}_j(\mathbf{c})} = p(\vec{k}'|\mathbf{c})k^{\rm out\prime}/\langle k\rangle
\end{aligned}
$$

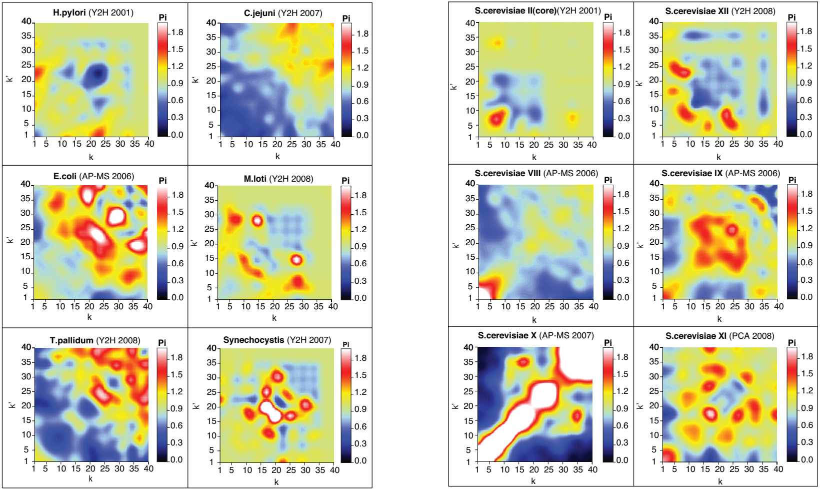
so focus on:

$$
\Pi(\vec{k},\vec{k}'|\mathbf{c}) = \frac{W(\vec{k},\vec{k}'|\mathbf{c})}{W_1(\vec{k}|\mathbf{c})W_2(\vec{k}'|\mathbf{c})}
$$

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

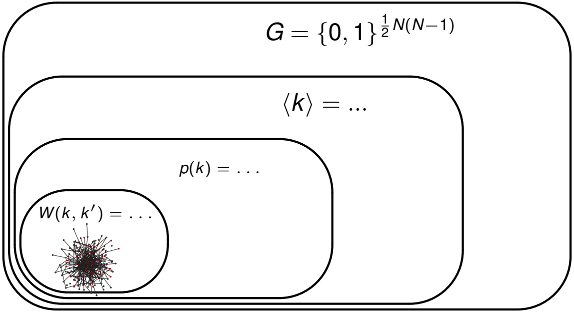
Information in degree correlations?

plot $\Pi(k, k') = W(k, k')/W(k)W(k')$
for protein interaction networks:



Graph classification
via increasingly detailed
feature prescription

e.g. nondirected graphs:



Tailored random graph ensembles

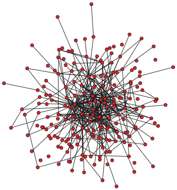
maximum entropy random graph ensembles,
$p(\mathbf{c})$ with prescribed values for $\langle k \rangle$, $p(k)$, $W(k,k'), \ldots$

- – proxies for real networks in stat mech models
- – complexity: how many networks exist with same features as $\mathbf{c}$? — counting
- – hypothesis testing: graphs with controlled features as null models — generating

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

random graph ensembles

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



- **Tailoring via hard constraints**

  (i) impose values for specific observables: $\Omega_\mu(\mathbf{c}) = \Omega_\mu$ for $\mu = 1 \ldots p$
  (ii) $p(\mathbf{c})$: all graphs that meet constraints are equally likely

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

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

note 1:

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

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

note 2:

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

- **Tailoring via soft constraints**

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

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

  parameters $\omega_\mu(\boldsymbol{\Omega})$:
  to be solved from

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

  note 1:

  all graphs $\mathbf{c}$ can in principle emerge; those with $\boldsymbol{\Omega}(\mathbf{c}) \approx \boldsymbol{\Omega}$ are the most likely

  note 2:

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

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

  note 3:

  for observables $\boldsymbol{\Omega}(\mathbf{c})$ that are *macroscopic in nature and* $\mathcal{O}(N^0)$,
  one will generally find deviations from $\boldsymbol{\Omega}(\mathbf{c}) = \boldsymbol{\Omega}$ to tend to zero as $N \to \infty$

Example 1a

nondirected graphs, $c_{ii}=0$ for all i,
impose average connectivity via hard 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 $\delta_{nm}=(2\pi)^{-1}\int_{-\pi}^{\pi}\mathrm{d}\omega\;\mathrm{e}^{\mathrm{i}(n-m)\omega}$

$$\begin{aligned}\mathcal{N}(\langle k\rangle) &= \int_{-\pi}^{\pi}\frac{\mathrm{d}\omega}{2\pi}\;\mathrm{e}^{\mathrm{i}\omega N\langle k\rangle}\sum_{\mathbf{c}}\mathrm{e}^{-\mathrm{i}\omega\sum_{ij}c_{ij}}=\int_{-\pi}^{\pi}\frac{\mathrm{d}\omega}{2\pi}\;\mathrm{e}^{\mathrm{i}\omega N\langle k\rangle}\prod_{i<j}\Big[\sum_{c_{ij}}\mathrm{e}^{-2\mathrm{i}\omega c_{ij}}\Big]\\
&= \int_{-\pi}^{\pi}\frac{\mathrm{d}\omega}{2\pi}\;\mathrm{e}^{\mathrm{i}\omega N\langle k\rangle}(1+\mathrm{e}^{-2\mathrm{i}\omega})^{\frac{1}{2}N(N-1)}\\
&= \sum_{\ell=0}^{\frac{1}{2}N(N-1)}\binom{\frac{1}{2}N(N-1)}{\ell}\int_{-\pi}^{\pi}\frac{\mathrm{d}\omega}{2\pi}\;\mathrm{e}^{\mathrm{i}\omega N\langle k\rangle-2\mathrm{i}\ell\omega}=\binom{\frac{1}{2}N(N-1)}{\frac{1}{2}N\langle k\rangle}\\
&= \mathrm{e}^{\frac{1}{2}N\langle k\rangle\left[\log(N/\langle k\rangle)+1\right]+\mathcal{O}(\log N)}\qquad \textit{Stirling}:\quad n!=\mathrm{e}^{n\log n-n+\mathcal{O}(\log n)}\quad n\to\infty\end{aligned}$$

Example 1b

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

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

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

$$\omega \textit{ solved from}: \qquad \langle k \rangle = \frac{1}{Z(\omega)} \sum_{\mathbf{c}} \Big(\frac{1}{N} \sum_{k\ell} c_{k\ell} \Big) \mathrm{e}^{\omega \sum_{ij} c_{ij}} = \frac{\mathrm{d}}{\mathrm{d}\omega} \frac{1}{N} \log Z(\omega)$$

- calculate $Z(\omega)$ and ω:

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

- Equivalently:

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

Example 2a

nondirected graphs, $c_{ii} = 0$ for all i,
impose degree sequence via <u>hard</u> constraint,
$\Omega_i(\mathbf{c}) = \sum_j c_{ij}, \quad i = 1 \ldots N$

- demand: $\sum_j c_{ij} = k_i$ for all i

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

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

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

$$\begin{aligned}
\mathcal{N}(\mathbf{k}) &= \int_{-\pi}^{\pi} \prod_i \left(\frac{\mathrm{d}\omega_i}{2\pi} \, \mathrm{e}^{\mathrm{i}\omega_i k_i} \right) \sum_{\mathbf{c}} \mathrm{e}^{-\mathrm{i} \sum_i \omega_i \sum_j c_{ij}} = \int_{-\pi}^{\pi} \frac{\mathrm{d}\boldsymbol{\omega} \mathrm{e}^{\mathrm{i}\boldsymbol{\omega}\cdot\mathbf{k}}}{(2\pi)^N} \prod_{i<j} \Big[\sum_{c_{ij}} \mathrm{e}^{-\mathrm{i}(\omega_i+\omega_j)c_{ij}} \Big] \\
&= \int_{-\pi}^{\pi} \frac{\mathrm{d}\boldsymbol{\omega} \, \mathrm{e}^{\mathrm{i}\boldsymbol{\omega}\cdot\mathbf{k}}}{(2\pi)^N} \prod_{i<j} (1 + \mathrm{e}^{-\mathrm{i}(\omega_i+\omega_j)}) = ?
\end{aligned}$$

possible (leading orders in N), but no longer obvious ...

Example 2b

nondirected graphs, $c_{ii} = 0$ for all i,
impose degree sequence via soft constraint,
$\Omega_i(\mathbf{c}) = \sum_j c_{ij}, \quad i = 1 \ldots N$

- demand: $\langle \sum_j c_{ij} \rangle = k_i$ for all i

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

$$\boldsymbol{\omega} \textit{ solved from}: \qquad \forall m: \quad k_m = \frac{1}{Z(\boldsymbol{\omega})} \sum_{\mathbf{c}} \Big(\sum_n c_{mn} \Big) \mathrm{e}^{\sum_i \omega_i \sum_j c_{ij}} = \frac{\partial}{\partial \omega_m} \log Z(\boldsymbol{\omega})$$

- calculate $Z(\boldsymbol{\omega})$ and $\boldsymbol{\omega}$:

$$\begin{aligned} k_m &= \frac{\partial}{\partial \omega_m} \log \sum_{\mathbf{c}} \mathrm{e}^{\sum_{i<j} c_{ij}(\omega_i+\omega_j)} = \frac{\partial}{\partial \omega_m} \log \prod_{i<j} \Big[\sum_{c_{ij}} \mathrm{e}^{c_{ij}(\omega_i+\omega_j)} \Big] \\ &= \sum_{i<j} \frac{\partial}{\partial \omega_m} \log(1+\mathrm{e}^{\omega_i+\omega_j}) = \frac{1}{2} \sum_{i \neq j} (\delta_{im}+\delta_{jm}) \frac{\mathrm{e}^{\omega_i+\omega_j}}{1+\mathrm{e}^{\omega_i+\omega_j}} = \sum_{i \neq m} \frac{\mathrm{e}^{\omega_i+\omega_m}}{1+\mathrm{e}^{\omega_i+\omega_m}} \end{aligned}$$

N transcendental eqns to be solved ...

Example 3a

nondirected graphs, $c_{ii} = 0$ for all i,
impose degree sequence and kernel $W(k,k')$ via hard constraint,

$\Omega_i(\mathbf{c}) = \sum_j c_{ij}, \quad i,j = 1\ldots N,$
$\Omega_{kk'}(\mathbf{c}) = \sum_{ij} c_{ij}\delta_{k,\sum_\ell c_{i\ell}}\delta_{k',\sum_\ell c_{j\ell}}, \quad k,k' \in \mathbb{N}$

- demand: $\sum_j c_{ij} = k_i$ for all i, and
  $\sum_{ij} c_{ij}\delta_{k,\sum_\ell c_{i\ell}}\delta_{k',\sum_\ell c_{j\ell}} = N\langle k\rangle W(k,k')$ for all (k,k')
  (with $\langle k\rangle = N^{-1}\sum_i k_i$)

$$p(\mathbf{c}|\mathbf{k},W) = \frac{\Big[\prod_i \delta_{\sum_j c_{ij},k_i}\Big]\Big[\prod_{k,k'}\delta_{\sum_{ij}c_{ij}\delta_{k,k_i}\delta_{k',k_j},N\langle k\rangle W(k,k')}\Big]}{\mathcal{N}(\mathbf{k},W)},$$

$$\mathcal{N}(\mathbf{k},W) = \sum_{\mathbf{c}}\Big[\prod_i \delta_{\sum_j c_{ij},k_i}\Big]\Big[\prod_{k,k'}\delta_{\sum_{ij}c_{ij}\delta_{k,k_i}\delta_{k',k_j},N\langle k\rangle W(k,k')}\Big]$$

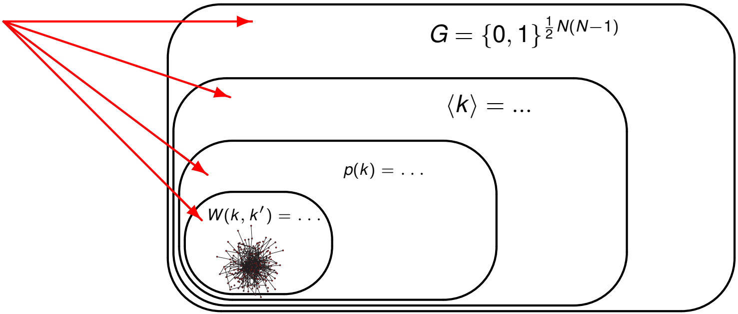
- calculate $\mathcal{N}(\mathbf{k},W)$:

$$\mathcal{N}(\mathbf{k},W) \;=\; \int_{-\pi}^{\pi}\prod_i\Big(\frac{\mathrm{d}\omega_i}{2\pi}\mathrm{e}^{\mathrm{i}\omega_i k_i}\Big)\Big(\prod_{k,k'}\frac{\mathrm{d}\psi_{kk'}}{2\pi}\mathrm{e}^{\mathrm{i}\psi_{kk'}N\langle k\rangle W(k,k')}\Big)$$
$$\times\sum_{\mathbf{c}}\mathrm{e}^{-\mathrm{i}\sum_i\omega_i\sum_j c_{ij}-\mathrm{i}\sum_{kk'}\psi_{kk'}\sum_{ij}c_{ij}\delta_{k,k_i}\delta_{k',k_j}}$$

doable, but increasingly complicated....

Counting tailored graphs

how many graphs in each family?



entropy and complexity

properties of Shannon entropy (information theory)

- **effective nr of graphs** in ensemble $p(\mathbf{c}|\star)$:
  ($\star$: imposed observables)

$$\mathcal{N}(\star)=e^{N\langle k\rangle S(\star)},\qquad S(\star)=-\frac{1}{N\langle k\rangle}\sum_{\mathbf{c}} p(\mathbf{c}|\star)\log p(\mathbf{c}|\star)\quad (\textit{entropy per link})$$

- $S(\star)$: proportional to the average nr of bits one needs to specify
  to identify a member graph $\mathbf{c}$ in the ensemble

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

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

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

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

  $\exists$ many graphs with feature $\star$: graphs with $\star$ have low complexity
  $\exists$ few graphs with feature $\star$: graphs with $\star$ have high complexity

Shannon entropy per bond
final result for nondirected graphs

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

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

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

$\pi(\ell) = \mathrm{e}^{-\langle k \rangle} \langle k \rangle^{\ell} / \ell!$
degree distr of Erdös-Renyi graphs

degree complexity: proportional to Kullback-Leibler distance (so ≥ 0)
wiring complexity: proportional to mutual information (so ≥ 0)

Shannon entropy per bond
final result for directed graphs:

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

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

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

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

$\pi(\ell) = {\rm e}^{-\langle k \rangle} \langle k \rangle^{\ell} / \ell!$
$\pi(k^{\rm in})\pi(k^{\rm out})$: degree distr of directed Erdös-Renyi graphs

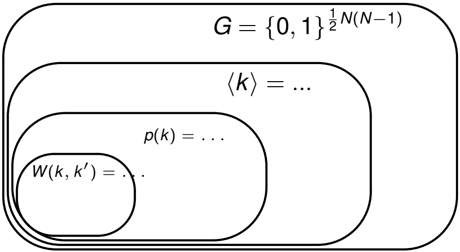
degree complexity: proportional to Kullback-Leibler distance (so ≥ 0)
wiring complexity: proportional to mutual information (so ≥ 0)

Generating tailored random graphs numerically

next:

generate
tailored random graphs

from these families
numerically ...



typical questions

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

how to generate

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

similar for directed graphs ...

why is the generation of graphs a nontrivial issue?

- many users underestimate/misjudge what the real problem is:

  sampling the space of all graphs with given features: usually easy ...
  sampling them with required probabilities: nontrivial!

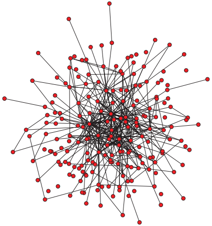
- many ad-hoc graph generation algorithms that *appear* sensible, but without proper analysis of which measure they converge to

- in cellular biology graphs are often used as 'null models', against which to test hypotheses on observed features in signalling networks

  if these null models are *biased*,
  the hypothesis test is fundamentally flawed ...

trivial case: no constraints
standard Glauber/MCMC dynamics

(Metropolis et al 1953)



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 the distribution $p(\mathbf{c})$

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

(trivially generalised
to directed graphs)

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 $\mathbf{k} = (k_1, \ldots, k_N)$

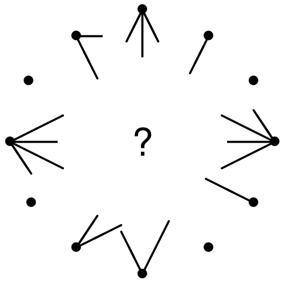
strategy: stochastic growth dynamics, starting from graph with no links

- 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
  $c_{ij} = 0 \;\rightarrow\; c_{ij} = 1$

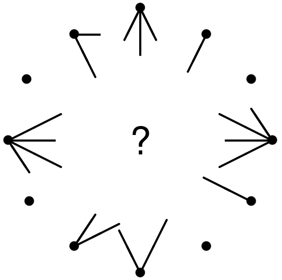
terminate if $\sum_j c_{ij} = k_i$ for all i



(trivially generalised to directed graphs)

Matching algorithm

limitations and problems ...



- major limitation:

  cannot control graph probabilities, just aims to generate $\mathbf{c} \in G[\mathbf{k}]$ with equal probs

- inconvenience: convergence not guaranteed

  process can 'hang' before $\sum_j c_{ij} = k_i$ for all i
  if one remaining 'stub' requires self-loops
  (happens more often when there are 'hubs',
  i.e. nodes with large degree)

  – monitor the evolving degrees, to test for this
  – if process 'hangs': reject and start over 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 ...

Edge switching algorithm

(Seidel, 1976)

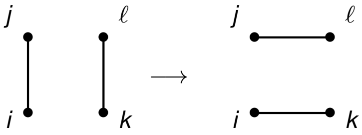
objective: generate random nondirected graph $\mathbf{c} \in \{0,1\}^{\frac{1}{2}N(N-1)}$ with specified degree sequence $\mathbf{k} = (k_1, \ldots, k_N)$

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), where $\mathbf{c}^0$ is some graph with the correct degrees

repeat:

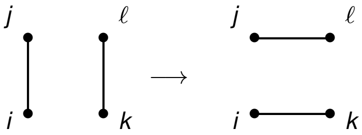
- pick at random four nodes (i,j,k,ℓ) that are *pairwise connected*
- carry out an 'edge swap' (or 'Seidel switch), see diagram (preserves all degrees!)



terminate if stochastic process has equilibrated

Edge switching algorithm

limitations and problems ...



- major limitation:

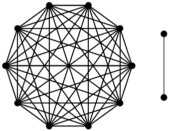
  cannot control graph probabilities, aims to generate $\mathbf{c} \in G[\mathbf{k}]$ with equal probs

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

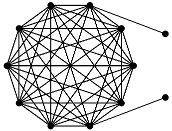
- sampling bias:

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

many possible moves



only one move ...



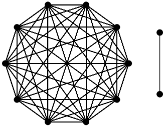
nr of possible moves depends on state **c**!

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

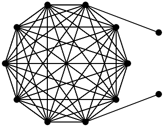
dangerous for scale-free graphs ...

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

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



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



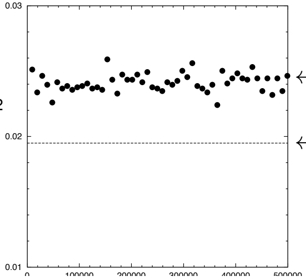
for flat measure:

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

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

'accept all'
edge swapping:

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



←*simulation*

←*theory*

executed moves

need to study graph dynamics more systematically ...

Monte Carlo processes for constrained graphs

- 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 the move $\mathbf{c}' \to \mathbf{c}$
  $n \in \mathbb{N}$: algorithmic time

$$\forall \mathbf{c} \in G[\star]: \qquad p_{n+1}(\mathbf{c}) = \sum_{\mathbf{c}' \in G[\star]} W(\mathbf{c}|\mathbf{c}')p_n(\mathbf{c}')$$

- 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 $G[\star]$

MCMC objective

construct transition probs $W(\mathbf{c}|\mathbf{c}')$, based on moves $F\in\Phi$,
such that process converges to $p(\mathbf{c})=Z^{-1}e^{-H(\mathbf{c})}$ on $G[\star]$

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

$q(F|\mathbf{c})$: *move proposal probability*

$A(\mathbf{c}|\mathbf{c}')$: *move acceptance probability*

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

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

- detailed balance condition:

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

if allowed F equally probable:
$q(F|\mathbf{c})=I_F(\mathbf{c})/n(\mathbf{c})$

$$(\forall F\in\Phi)(\forall \mathbf{c}\in G_F[\star]):\qquad \frac{1}{n(\mathbf{c})}A(F\mathbf{c}|\mathbf{c})\mathrm{e}^{-H(\mathbf{c})}=\frac{1}{n(F\mathbf{c})}A(\mathbf{c}|F\mathbf{c})\mathrm{e}^{-H(F\mathbf{c})}$$

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

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

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

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

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

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

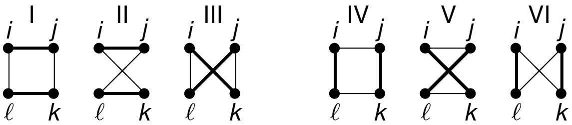
bookkeeping of elementary moves

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

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

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

  possible edge swaps to act on (i,j,k,ℓ):



- group into pairs (I,IV), (II,V), and (III,VI)
  auto-invertible swaps: $F_{ijk\ell;\alpha}$, with $i<j<k<\ell$ and $\alpha \in \{1,2,3\}$

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

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

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

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

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

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

combinatorial problem:
$(\overline{\delta}_{ij} = 1-\delta_{ij})$

$$
\begin{aligned}
n(\mathbf{c}) &= \sum_{i<j<k<\ell} \overbrace{\Big(I_{ijk\ell;1(\mathbf{c})} + I_{ijk\ell;2(\mathbf{c})} + I_{ijk\ell;3(\mathbf{c})}\Big)}^{\textit{invariant under all permutations of } (i,j,k,\ell)} \\
&= \frac{1}{4!}\sum_{ijk\ell} \overline{\delta}_{ij}\overline{\delta}_{ik}\overline{\delta}_{i\ell}\overline{\delta}_{jk}\overline{\delta}_{j\ell}\overline{\delta}_{k\ell} \sum_{\alpha=1}^{3} I_{ijk\ell;\alpha(\mathbf{c})} \qquad (\textit{permutation invariance}) \\
&= \frac{1}{4}\sum_{ijk\ell} \overline{\delta}_{ij}\overline{\delta}_{ik}\overline{\delta}_{i\ell}\overline{\delta}_{jk}\overline{\delta}_{j\ell}\overline{\delta}_{k\ell} c_{ij}c_{k\ell}(1-c_{i\ell})(1-c_{jk}) \qquad (\textit{permutation, inversion}) \\
&= \frac{1}{4}\sum_{ijk\ell} \overline{\delta}_{ik}\overline{\delta}_{i\ell}\overline{\delta}_{jk}\overline{\delta}_{j\ell} c_{ij}c_{k\ell}(1-c_{i\ell})(1-c_{jk}) \qquad (\textit{no diagonal entries})
\end{aligned}
$$

work out remaining terms explicitly ...

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

Examples:

- Fully connected graphs:
  $k_i = N-1$ for all i, $\mathrm{Tr}(\mathbf{c}^4) = (N-1)[(N-1)^3+1]$, $\mathrm{Tr}(\mathbf{c}^3) = N(N-1)(N-2)$
  formula: $n(\mathbf{c}) = 0$ *(ok by inspection)*

- Periodic chains $c_{ij} = \delta_{i,j-1} + \delta_{i,j+1}$ (mod N), $N \geq 4$:
  $k_i = 2$ for all i, $\mathrm{Tr}(\mathbf{c}^4) = 6N$, $\mathrm{Tr}(\mathbf{c}^3) = 0$
  formula: $n(\mathbf{c}) = N(N-4)$ *(ok by inspection)*

- Two isolated links $c_{12} = c_{21} = c_{34} = c_{43} = 1$, all other $c_{ij} = 0$:
  $k_1 = k_2 = k_3 = k_4 = 1$, $k_{i>4} = 0$, $\mathrm{Tr}(\mathbf{c}^4) = 4$, $\mathrm{Tr}(\mathbf{c}^3) = 0$
  formula: $n(\mathbf{c}) = 2$ *(ok by inspection)*

- Regular random graphs with $p(k) = \delta_{k,2}$:
  use eigenvalue distribution of $\mathbf{c}$ (Dorogovtsev 2003),
  formula: $n(\mathbf{c}) = N(N-4) + o(N)$

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

practicalities

how to avoid calculating $n(\mathbf{c})$ at each iteration step,

- use simple bounds:

$$\frac{N}{4}\Big(N\langle k\rangle^2+\langle k\rangle-\langle k^2\rangle\Big) - \frac{N}{2}\langle k^2\rangle k_{\max} \leq n(\mathbf{c}) \leq \frac{N}{4}\Big(N\langle k\rangle^2+\langle k\rangle-\langle k^2\rangle\Big)$$

  state-dependent part can be ignored if $\langle k^2\rangle k_{\max}/\langle k\rangle^2 \ll N$

- (i) calculate $n(\mathbf{c})$ only at time $n=0$

  (ii) update $n(\mathbf{c})$ dynamically, by calculating at each step change $\Delta_{ijk\ell;\alpha} n(\mathbf{c})$ for executed move $F_{ijk\ell;\alpha}$

  e.g.

$$\Delta_{ijk\ell;\alpha}\mathrm{Tr}(\mathbf{c}^3) = 6 \sum_{(a,b)\in S_{ijk\ell;\alpha},\ a<b} (1-2c_{ab}) \sum_{v\notin\{i,j,k,\ell\}} c_{bv}c_{va}$$

$$\Delta_{ijk\ell;\alpha}\mathrm{Tr}(\mathbf{c}^4) = \textit{ more complicated but explicit formula ...}$$

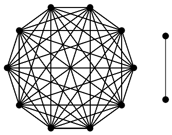
Example:

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

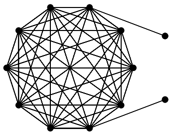
$N = 100$

naive versus correct
acceptance probabilities

many possible moves



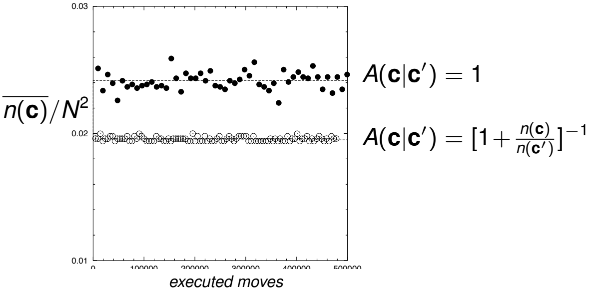
only one move ...



predictions:

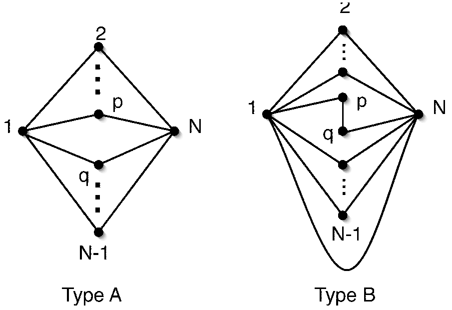
$p(\mathbf{c}) = \textit{constant}$:
$\overline{n(\mathbf{c})}/N^2 \approx 0.0195$

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

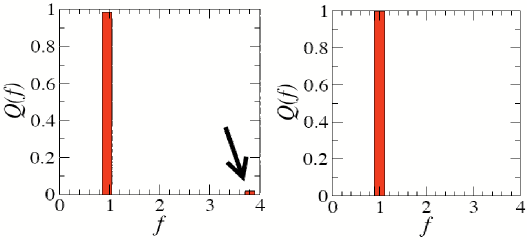


Example

graph type A: $n(\mathbf{c}) = K(K-1)$
graph type B: $n(\mathbf{c}) = 2(K-1)$



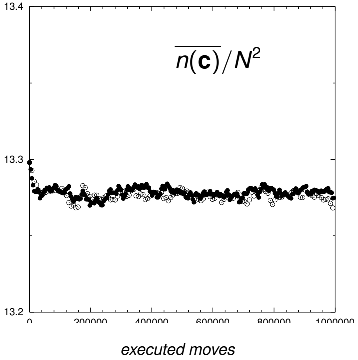
measure distribution $Q(f)$ of (rescaled) frequencies at which graphs are visited

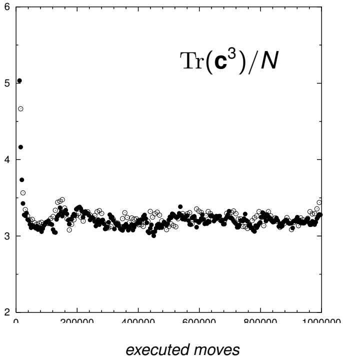


'accept all' *correct dynamics*

Example

human protein interaction network
$N = 9463$, $\langle k \rangle \approx 7.4$



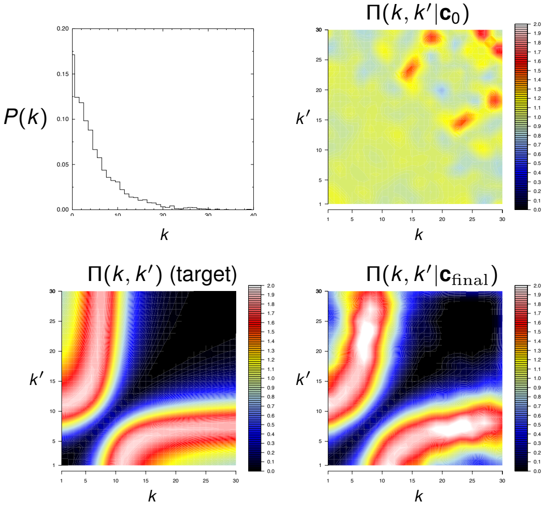


●: 'accept all' edge swap dynamics
o: correct edge swap dynamics

(so no serious harm done yet ...)

Example

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



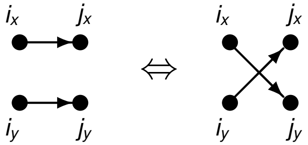
$N = 4000$,
$\overline{k} = 5$

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

bookkeeping of elementary moves

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

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



- auto-invertible edge-swaps:
  Let $\Lambda = \{(i,j)\in N^2|\ c_{ji}=1\}$

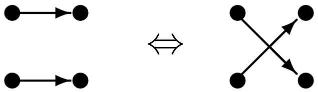
$$I_{(i_x,j_x),(i_y,j_y);\Box} = \begin{cases} 1 & \textit{if } (i_x,j_x),(i_y,j_y)\in\Lambda \text{ and } (i_x,j_y),(i_y,j_x)\notin\Lambda \\ 0 & \text{otherwise}\end{cases}$$

If $I_{(i_x,j_x),(i_y,j_y);\Box}=1$:

$$\begin{aligned} F_{(i_x,j_x),(i_y,j_y);\Box}(\mathbf{c})_{ij} &= 1-c_{ij} && \text{if } i\in\{i_x,i_y\} \text{ and } j\in\{j_x,j_y\} \\ F_{(i_x,j_x),(i_y,j_y);\Box}(\mathbf{c})_{ij} &= c_{ij} && \text{otherwise}\end{aligned}$$

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

for **directed** graphs:
are edge swaps *ergodic* set of moves?



Rao, 1996:

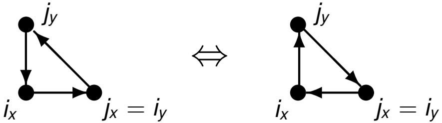
unless self-interactions are allowed,
edge swaps *not ergodic for directed graphs*

proof:
by counterexample

these two $N = 3$ graphs
are both in $G[\mathbf{k}^{\rm in}, \mathbf{k}^{\rm out}]$,
with $\mathbf{k}^{\rm in} = \mathbf{k}^{\rm out} = (1, 1, 1)$

but *no edge swap maps one to the other*

further move type required
to restore ergodicity:
3-loop reversal



$$I_{(i_x,j_x),(i_y,j_y);\triangle} = \begin{cases} 1 & \textit{if } (i_x,j_x),(i_y,j_y),(j_y,i_x) \in \Lambda \ \textit{ and } \ x_j = y_i \\ & \textit{and } \ (j_x,i_x),(j_y,i_y),(i_x,j_y) \notin \Lambda \\ 0 & \text{otherwise} \end{cases}$$

$$\begin{aligned} F_{(i_x,j_x),(i_y,j_y);\triangle}(\mathbf{c})_{ij} &= 1 - c_{ij} \quad \textit{for } (i,j) \in \mathcal{S}_{i_x,j_x,j_y} \\ F_{(i_x,j_x),(i_y,j_y);\triangle}(\mathbf{c})_{ij} &= c_{ij} \qquad\quad \textit{for } (i,j) \notin \mathcal{S}_{i_x,j_x,j_y} \end{aligned}$$

$$\mathcal{S}_{abc} = \{(a,b),(b,c),(c,a),(b,a),(c,b),(a,c)\}$$

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

$$
\begin{aligned}
n(\mathbf{c}) &= n_{\Box}(\mathbf{c}) + n_{\triangle}(\mathbf{c}) \\
&= \sum_{(i_x,j_x),(i_y,j_y)\in\Lambda} I_{(i_x,j_x),(i_y,j_y);\Box} + \sum_{(i_x,j_x),(i_y,j_y)\in\Lambda} I_{(i_x,j_x),(i_y,j_y);\triangle}
\end{aligned}
$$

$$
\begin{aligned}
I_{(i_x,j_x),(i_y,j_y);\Box} &= c_{i_x,j_x} c_{i_y,j_y} (1-c_{i_x,j_y})(1-c_{i_y,j_x}) \\
I_{(i_x,j_x),(i_y,j_y);\triangle} &= \delta_{x_j,y_i}\; c_{i_x,j_x} c_{i_y,j_y} c_{j_y,i_x} (1-c_{j_x,i_x})(1-c_{j_y,i_y})(1-c_{i_x,j_y})
\end{aligned}
$$

combinatorial problem again easily solved:

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

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

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

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

$$n_{\triangle}(\mathbf{c}) = \frac{1}{3}{\rm Tr}(\mathbf{c}^3) - {\rm Tr}(\hat{\mathbf{c}}\mathbf{c}^2) + {\rm Tr}(\hat{\mathbf{c}}^2\mathbf{c}) - \frac{1}{3}{\rm Tr}(\hat{\mathbf{c}}^3)$$

practicalities

how to avoid calculating $n_{\Box}(\mathbf{c})$ and $n_{\triangle}(\mathbf{c})$ at each iteration step,

- use simple bounds on $n_{\Box}(\mathbf{c})$ and $n_{\triangle}(\mathbf{c})$,
  state-dependent part can be ignored if

$$\frac{1}{\langle k\rangle} + \frac{2}{\langle k\rangle^2}\Big(k_{\rm max}^{\rm in}\langle k^{{\rm out}\ 2}\rangle + k_{\rm max}^{\rm out}\langle k^{{\rm in}\ 2}\rangle\Big) \ll N$$

- (i) calculate $n_{\Box}(\mathbf{c})$ and $n_{\triangle}(\mathbf{c})$ only at time $n=0$
  (ii) update $n_{\Box}(\mathbf{c})$ and $n_{\triangle}(\mathbf{c})$ dynamically, by calculating at each step
  change $\Delta_{ijk\ell;\alpha}n_{\Box}(\mathbf{c})$ and $\Delta_{ijk\ell;\alpha}n_{\triangle}(\mathbf{c})$ for executed move $F_{ijk\ell;\alpha}$

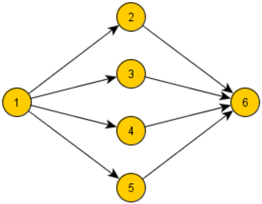
Example

$(k_1^{\mathrm{in}}, k_1^{\mathrm{out}}) = (0, N-2)$

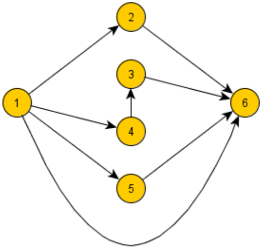
$i = 2 \ldots N-1$:
$(k_i^{\mathrm{in}}, k_i^{\mathrm{out}}) = (1, 1)$

$(k_N^{\mathrm{in}}, k_N^{\mathrm{out}}) = (N-2, 0)$

$(N-2)(N-3)$ moves



$2N-7$ moves



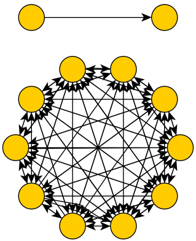
predicted values versus
equilibrated dynamics for $\overline{n(\mathbf{c})}/N^2$:

	prediction for $p(\mathbf{c}) = const$	dynamics with $A(\mathbf{c}\vert\mathbf{c}') = 1$	dynamics with $A(\mathbf{c}\vert\mathbf{c}') = [1+\frac{n(\mathbf{c})}{n(\mathbf{c}')}]^{-1}$
$N = 17$:	27.87	33.59	27.87
$N = 27$:	47.92	58.32	47.95

Example

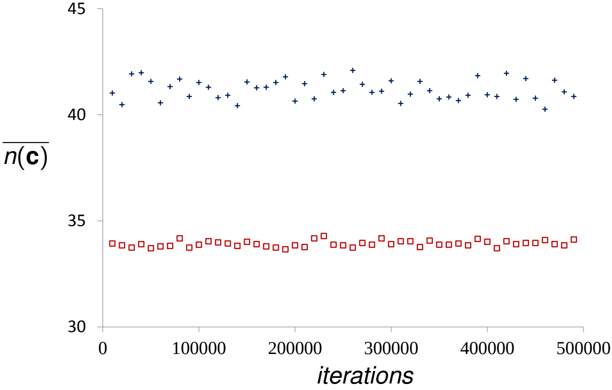
fully connected 'core' of $N-2$ nodes, plus two extra nodes

$N = 20$, target: flat measure



'accept all' edge swapping:
$\overline{n(\mathbf{c})}\rangle \approx 41.09$
predicted: 41.03

edge swapping with correct acceptance probabilities:
$\overline{n(\mathbf{c})} \approx 33.92$
predicted: 33.89



some references

network reviews

R Albert and AL Barabasi, Rev Mod Phys 74, 2002
SN Dorogovtsev, AV Goltsev, and JFF Mendes, Rev Mod Phys 80, 2008
ACC Coolen, F Fraternali, A Annibale, L Fernandes and J Kleinjung, in Handbook of Statistical Systems Biology, Wiley, 2011

entropies of tailored graph ensembles

G Bianconi, ACC Coolen, CJ Perez Vicente, Phys Rev E 78, 2008
A Annibale, ACC Coolen, L P Fernandes, F Fraternali, J Kleinjung, J Phys A 42, 2009
ES Roberts, T Schlitt and ACC Coolen, J Phys A 44, 2011
ES Roberts and ACC Coolen, J Phys A 47, 2014

graph dynamics and generation

R Taylor, in Combinatorial Mathematics VIII, Springer Lect Notes Math 884, 1981
AR Rao, R Jana, and S Bandyopadhya, Indian J of Statistics 58, 1996
ACC Coolen, A De Martino and A Annibale, J Stat Phys 136, 2009
K Roberts and ACC Coolen, Phys Rev E 85, 2012

website

www.mth.kcl.ac.uk/$\sim$tcoolen