Modelling of Complex Real-World Systems
Part A. General Methods

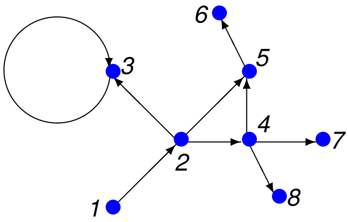
A3. Networks and Graphs

Ton Coolen
Department of Biophysics, Radboud University
module NWI-NM127, January 2021

Definitions

Networks and Graphs

- *N-node graph* $G(V,E)$: set of vertices/nodes $V = \{1,\ldots,N\}$
  set of edges/links $E \subseteq \{(i,j)|\ i,j \in V\}$
- *simple graph*: no self-links, $\forall (i,j) \in E:\ i \neq j$
- *nondirected graph*: symmetric links only, if $(i,j) \in E$ then $(j,i) \in E$
- *directed graph*: contains non-symmetric links, $\exists (i,j) \in E$ with $(j,i) \notin E$



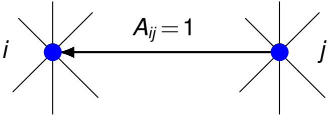
$$V = \{1,2,3,4,5,6,7,8\}$$

$$E = \{(2,1),(3,2),(4,2),(5,2),(3,3),(5,4),(7,4),(8,4),(6,5)\}$$

Adjacency matrix

adjacency matrix $\mathbf{A} \in \{0,1\}^{N\times N}$:
fully equivalent definition of graphs

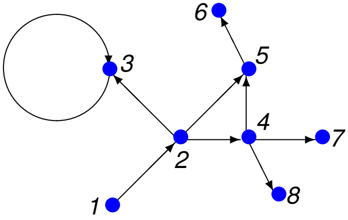
$$\forall (i,j): \quad \begin{array}{llll} A_{ij} = 1 & \text{if} & (i,j) \in E, & \text{i.e. there is a link } j \to i \\ A_{ij} = 0 & \text{if} & (i,j) \notin E, & \text{i.e. there is no link } j \to i \end{array}$$



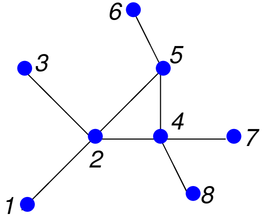
- *simple graph* $\Leftrightarrow$ $A_{ii} = 0$ for all i
- *nondirected graph* $\Leftrightarrow$ $A_{ij} = A_{ji}$ for all (i,j)
- *directed graph* $\Leftrightarrow$ $A_{ij} \neq A_{ji}$ for some (i,j)

advantage:
convert graph analyses to matrix manipulations

examples



$$\Leftrightarrow \quad \mathbf{A} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}$$



$$\Leftrightarrow \quad \mathbf{A} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Paths in networks

$$\prod_{\ell=1}^{k-1} A_{i_\ell i_{\ell+1}} = 1 \qquad \text{if the graph contains the path of connected links } i_k \to i_{k-1} \to \ldots \to i_2 \to i_1$$

$$\prod_{\ell=1}^{k-1} A_{i_\ell i_{\ell+1}} = 0 \qquad \text{if it does not}$$

$$\sum_{i_1=1}^{N} \cdots \sum_{i_k=1}^{N} A_{ii_1} \Big(\prod_{\ell=1}^{k-1} A_{i_\ell i_{\ell+1}} \Big) A_{i_k j} > 0 \quad \Leftrightarrow \quad \text{there exists a path of length } k+1 \text{ from node } j \text{ to node } i$$

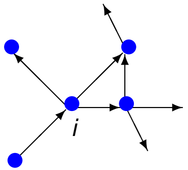
$$\sum_{i_1=1}^{N} \cdots \sum_{i_k=1}^{N} A_{ii_1} \Big(\prod_{\ell=1}^{k-1} A_{i_\ell i_{\ell+1}} \Big) A_{i_k j} = 0 \quad \Leftrightarrow \quad \text{there exists no path of length } k+1 \text{ from node } j \text{ to node } i$$

$$(\mathbf{A}^{k+1})_{ij} > 0 \quad \Leftrightarrow \quad \text{there exists a path of length } k+1 \text{ from node } j \text{ to node } i$$

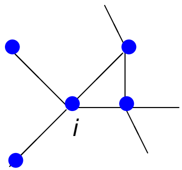
$$(\mathbf{A}^{k+1})_{ij} = 0 \quad \Leftrightarrow \quad \text{there exists no path of length } k+1 \text{ from node } j \text{ to node } i$$

Node degrees

- *in-degree of node i*: $k_i^{\mathrm{in}} = \sum_{j=1}^{N} A_{ij}$
- *out-degree of node i*: $k_i^{\mathrm{out}} = \sum_{j=1}^{N} A_{ji}$
- *non-directed graphs*: $k_i^{\mathrm{in}} = k_i^{\mathrm{out}} = k_i$ for all i
- *degree sequence in non-directed graphs*: $\mathbf{k} = (k_1, k_2, \ldots, k_N)$



$k_i^{\mathrm{in}} = 1, \; k_i^{\mathrm{out}} = 3$



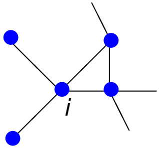
$k_i = 4$

Clustering and closed paths

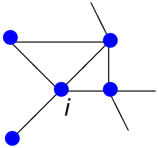
- *clustering coefficient* $C_i(\mathbf{A})$ of node i with degree ≥ 2 (nondirected graphs only)

$$C_i(\mathbf{A}) = \frac{\text{number of } \textit{connected} \text{ node pairs among neighbours of } i}{\text{number of node pairs among neighbours of } i}$$

$$= \frac{\sum_{j,k=1}^{N}(1-\delta_{jk})A_{ij}A_{jk}A_{ik}}{\sum_{j,k=1}^{N}(1-\delta_{jk})A_{ij}A_{ik}} \in [0,1]$$



$C_i = 1/6$



$C_i = 2/6 = 1/3$

- *number of closed paths of length* $\ell > 0$

$$L_\ell(\mathbf{A}) = \sum_{i_1=1}^{N} \dots \sum_{i_\ell=1}^{N} \Big(\prod_{k=1}^{\ell-1} A_{i_k,i_{k+1}}\Big) A_{i_\ell,i_1} = \sum_{i=1}^{N} (\mathbf{A}^\ell)_{ii} = \mathrm{Tr}(\mathbf{A}^\ell)$$

Macroscopic structure

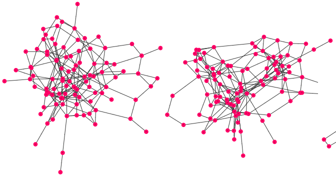
Averages and distributions of single-node quantities

- *average in-degree*: $\bar{k}^{\rm in} = N^{-1}\sum_{i=1}^{N} k_i^{\rm in}$
- *average out-degree*: $\bar{k}^{\rm out} = N^{-1}\sum_{i=1}^{N} k_i^{\rm out}$
- *average clustering coefficient*: $\bar{C} = N^{-1}\sum_{i=1}^{N} C_i(\mathbf{A})$
  ($C_i \equiv 0$ if $k_i < 2$)

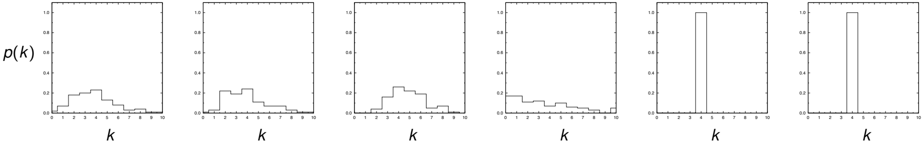
- *degree distribution of non-directed graph*
  $$\forall k \in \mathbb{N}: \quad p(k) = \frac{1}{N}\sum_i \delta_{k,k_i}$$

- *joint in- and out degree distribution of directed graph*
  $$\forall (k^{\rm in}, k^{\rm out}) \in \mathbb{N}^2: \quad p(k^{\rm in}, k^{\rm out}) = \frac{1}{N}\sum_i \delta_{k^{\rm in},k_i^{\rm in}}\delta_{k^{\rm out},k_i^{\rm out}}$$

examples with $N = 100$
(non-directed)



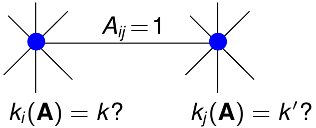
Erdös-Rényi *Modular* *Small world* *Scale-free* *Regular random* *Periodic lattice*



Distributions of multi-node quantities

- *joint distribution of degrees of connected node pairs* in non-directed graph

$$\forall k, k' \geq 0: \qquad W(k,k') = \frac{\sum_{i\neq j} \delta_{k,k_i(\mathbf{A})} A_{ij} \delta_{k',k_j(\mathbf{A})}}{\sum_{i\neq j} A_{ij}}$$



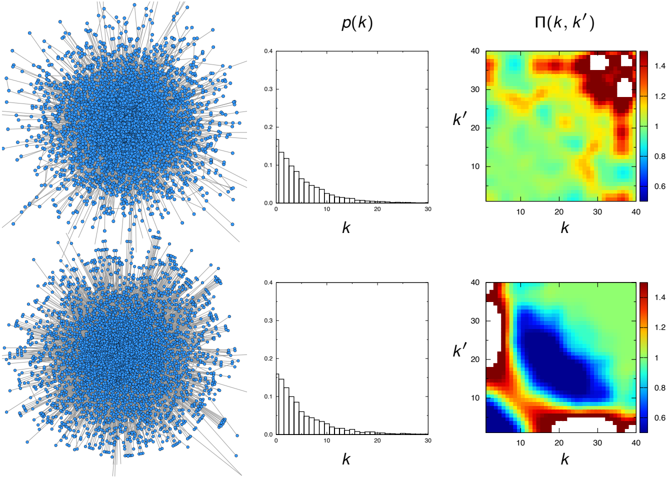
fraction of non-self links that connect a node of degree k to a node of degree k'

- *degree correlation ratio*

$$\Pi(k,k') = \frac{W(k,k')}{W(k)W(k')} = \frac{\bar{k}^2}{kk'} \frac{W(k,k')}{p(k)p(k')}$$

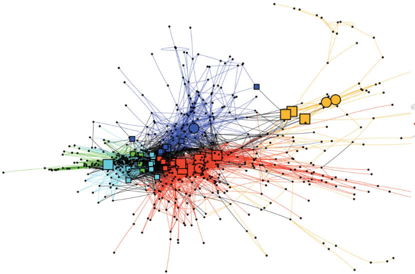
$\Pi(k,k') \neq 1$: degree correlations

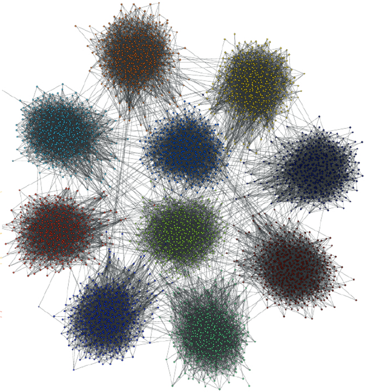
examples with $N = 5000$
(non-directed)



Modularity

nodes i have 'features' x_i,
nodes with same feature
more likely to connect





- *modularity* of non-directed graph

  $Q \in [-\frac{1}{2}, \frac{1}{2}]$
  $Q > 0$: modular graph
  $Q < 0$: anti-modular graph

$$Q = \frac{1}{2N\bar{k}} \sum_{ij} \left(A_{ij} - \frac{k_i k_j}{N\bar{k}} \right) \delta_{x_i, x_j}$$

(see exercises)

Eigenvalue spectra of adjacency matrices

for non-directed graphs

- *eigenvalue spectrum* of non-directed graph with eigenvalues $\{\mu_k\}$

$$\forall \mu \in \mathrm{I\!R}: \quad \varrho(\mu) = \frac{1}{N}\sum_{k=1}^{N} \delta[\mu - \mu_k]$$

- bounds: $\mu_{\min} \leq \bar{k} \leq \mu_{\max} \leq \max_j k_j$

- simple non-directed graphs

$$\int \mathrm{d}\mu\ \mu\varrho(\mu) = 0, \qquad \int \mathrm{d}\mu\ \mu^2\varrho(\mu) = \bar{k}, \qquad \int \mathrm{d}\mu\ \mu^\ell\varrho(\mu) = \frac{1}{N}L_\ell$$

L_ℓ: *nr of closed paths of length* ℓ

- if $\varrho(-\mu) = \varrho(\mu)$ for all $\mu \in \mathrm{I\!R}$, then graph has no closed paths of odd length (consequence of last identity)

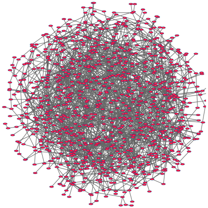
- proof:

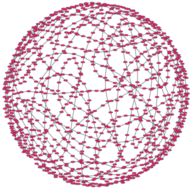
  transform **A** to diagonal form via unitary transformation, $\mathbf{A} = \mathbf{U}\mathbf{D}(\boldsymbol{\mu})\mathbf{U}^\dagger$, where $D(\boldsymbol{\mu})_{ij} = \mu_i \delta_{ij}$

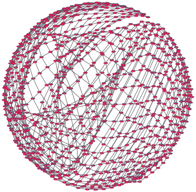
$$\begin{aligned}
N\int \mathrm{d}\mu\ \mu^\ell \varrho(\mu) &= \sum_k \mu_k^\ell = \sum_k \left[\mathbf{D}^\ell(\boldsymbol{\mu})\right]_{kk} = \sum_k \left[(\mathbf{U}^\dagger \mathbf{A}\mathbf{U})^\ell\right]_{kk} \\
&= \sum_k \left[\mathbf{U}^\dagger \left(\mathbf{A}\mathbf{U}\mathbf{U}^\dagger\right)^{\ell-1} \mathbf{A}\mathbf{U}\right]_{kk} = \sum_k \left[\mathbf{U}^\dagger \mathbf{A}^{\ell-1}\mathbf{A}\mathbf{U}\right]_{kk} \\
&= \sum_k \left[\mathbf{U}^\dagger \mathbf{A}^\ell \mathbf{U}\right]_{kk} = \sum_k \sum_{ij} (\mathbf{U}^\dagger)_{ki} (\mathbf{A}^\ell)_{ij} U_{jk} \\
&= \sum_k \sum_{ij} (\mathbf{A}^\ell)_{ij} U_{jk} (\mathbf{U}^\dagger)_{ki} = \sum_{ij} (\mathbf{A}^\ell)_{ij} (\mathbf{U}\mathbf{U}^\dagger)_{ji} = \sum_i (\mathbf{A}^\ell)_{ii}
\end{aligned}$$

$\ell = 1$: $\int \mathrm{d}\mu\ \mu^\ell \varrho(\mu) = \frac{1}{N}\sum_i A_{ii} = 0$
$\ell = 2$: $\int \mathrm{d}\mu\ \mu^2 \varrho(\mu) = \frac{1}{N}\sum_{ij} A_{ij}A_{ji} = \frac{1}{N}\sum_{ij} A_{ij} = \frac{1}{N}\sum_i k_i(\mathbf{A}) = \bar{k}$
general ℓ: use $\mathrm{Tr}(\mathbf{A}^\ell) = L_\ell$

examples with $N = 1000$
(non-directed)





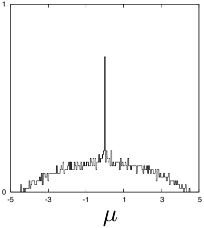


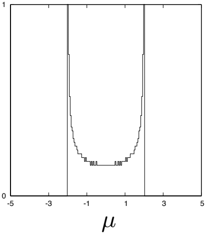
$\bar{k} = 4$
Erdös-Rényi

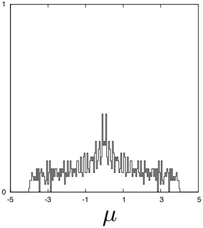
$\bar{k} = 2$
periodic ring

$\bar{k} = 4$
periodic 2D lattice

$\varrho(\mu)$







Definitions

Definitions

- *random graph ensemble* $\{\Omega, p\}$:

  (i) set Ω of adjacency matrices $\mathbf{A}$,
  (ii) probabilities $p(\mathbf{A})$ for each $\mathbf{A} \in \Omega$

  ensemble averages
  $$\langle f \rangle = \sum_{\mathbf{A} \in \Omega} p(\mathbf{A}) f(\mathbf{A})$$

- *Erdös-Rényi model*:
  simple nondirected graphs, i.i.d. links

  $$\Omega = \{\mathbf{A} \in \{0,1\}^{N \times N} |\ A_{ij} = A_{ji} \text{ and } A_{ii} = 0\ \forall (i,j)\}$$
  $$p(\mathbf{A}) = \prod_{i<j=1}^{N} p(A_{ij}), \qquad p(A_{ij}) = p^{\star} \delta_{A_{ij},1} + (1 - p^{\star}) \delta_{A_{ij},0}$$

  now two types of averages, e.g.
  $\bar{k} = \bar{k}(\mathbf{A}) = N^{-1} \sum_i k_i(\mathbf{A})$ vs $\langle k \rangle = \sum_{\mathbf{A} \in \Omega} p(\mathbf{A}) \bar{k}(\mathbf{A})$

Properties of ER model

- *average value of average degree*

$$\begin{aligned}\langle\bar{k}(\mathbf{A})\rangle &= \sum_{\mathbf{A}\in\Omega} p(\mathbf{A})\frac{1}{N}\sum_{rs} A_{rs} = \sum_{\mathbf{A}\in\Omega} p(\mathbf{A})\frac{2}{N}\sum_{r<s} A_{rs} = \frac{2}{N}\sum_{r<s}\sum_{A_{rs}} p(A_{rs})A_{rs} \\ &= \frac{2}{N}\frac{1}{2}N(N-1)\sum_{A\in\{0,1\}}(p^\star\delta_{A1}+(1-p^\star)\delta_{A0})A = (N-1)p^\star\end{aligned}$$

- *graphs* **A** *with same number* $L(\mathbf{A})$ *of links are equally probable*

$$\begin{aligned}p(\mathbf{A}) &= \prod_{i<j}\left[(p^\star)^{A_{ij}}(1-p^\star)^{1-A_{ij}}\right] = (p^\star)^{\sum_{i<j}A_{ij}}(1-p^\star)^{\frac{1}{2}N(N-1)-\sum_{i<j}A_{ij}} \\ &= (p^\star)^{L(\mathbf{A})}(1-p^\star)^{N(N-1)/2-L(\mathbf{A})}\end{aligned}$$

- *probabilities can be written as*

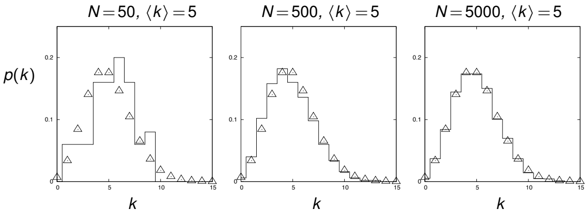
$$p(A_{ij}) = \frac{\langle k\rangle}{N-1}\delta_{A_{ij},1} + (1-\frac{\langle k\rangle}{N-1})\delta_{A_{ij},0}$$

ER model in the finite connectivity regime

$N \to \infty$ with $\langle k \rangle$ finite
$p^\star = \mathcal{O}(N^{-1})$

- *degree distribution: Poisson*

$$\lim_{N\to\infty} \langle p(k|\mathbf{A})\rangle = \mathrm{e}^{-\langle k\rangle}\langle k\rangle^k / k!$$

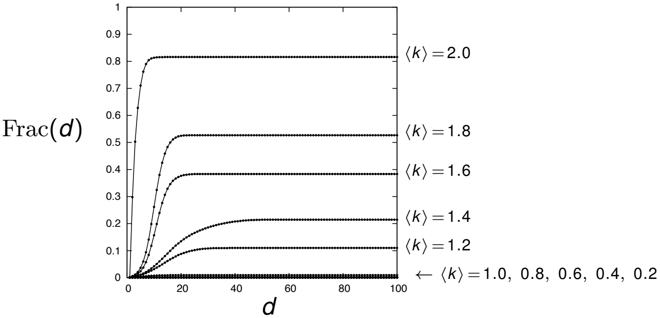


- *average clustering coefficient*

$$\langle C_i(\mathbf{A})\rangle = \frac{\langle k\rangle}{N}\left[1 - \mathrm{e}^{-\langle k\rangle} - \langle k\rangle \mathrm{e}^{-\langle k\rangle}\right] + \mathcal{O}(N^{-2})$$

Percolation phase transition in random tree-like graphs

giant component: finite fraction of nodes mutually connected
emerges at $\langle k \rangle = 1$, *percolation phase transition*



d_{ij}: distance in graph between nodes *i* and *j*

$\mathrm{Frac}(d) = \frac{2}{N(N-1)} \sum_{i<j} I[d_{ij} \leq d]$