Protein interaction networks via Tailored graph ensembles: a study of sampling

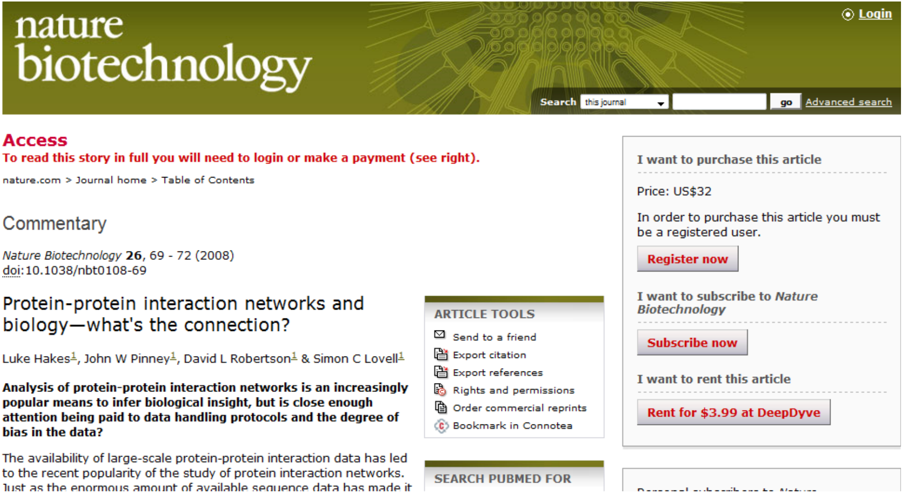
A Annibale

Department of Mathematics
King's College London

Outline

1. Motivation
2. Quantifying biases
   - Tailored random graph ensembles
   - Sampling protocols
   - Results
3. Inferring the true network from imperfect data
   - Bayesian analysis

Motivation



Access

To read this story in full you will need to login or make a payment (see right).

nature.com > Journal home > Table of Contents

Commentary

Nature Biotechnology **26**, 69 - 72 (2008)
doi:10.1038/nbt0108-69

Protein-protein interaction networks and biology—what's the connection?

Luke Hakes[1], John W Pinney[1], David L Robertson[1] & Simon C Lovell[1]

Analysis of protein-protein interaction networks is an increasingly popular means to infer biological insight, but is close enough attention being paid to data handling protocols and the degree of bias in the data?

The availability of large-scale protein-protein interaction data has led to the recent popularity of the study of protein interaction networks. Just as the enormous amount of available sequence data has made it

I want to purchase this article

Price: US$32

In order to purchase this article you must be a registered user.

Register now

I want to subscribe to *Nature Biotechnology*

Subscribe now

I want to rent this article

Rent for $3.99 at DeepDyve

Personal subscribers to Nature

- Large-scale PIN dataset are available nowdays
- Data are biased and incomplete

Is it possible to use the available data reliably?

Yes, if we understand the relation between the patterns of a real graph and those of a graph sample

Outline

1. Motivation
2. Quantifying biases
   - Tailored random graph ensembles
   - Sampling protocols
   - Results
3. Inferring the true network from imperfect data
   - Bayesian analysis

Graphs/Networks in a nutshell

- size N; 'nodes': $i, j, k = 1...N$

Graphs/Networks in a nutshell

- size N; 'nodes': $i, j, k = 1...N$
- 'links': Connectivity matrix

$$c_{ij} = \begin{cases} 1 & i—j \\ 0 & otherwise \end{cases}$$

Graphs/Networks in a nutshell

- size N; 'nodes': $i, j, k = 1...N$
- 'links': Connectivity matrix $$c_{ij} = \begin{cases} 1 & i—j \\ 0 & otherwise \end{cases}$$
- degrees: $k_i = \sum_j c_{ij}$; $\mathbf{k} = (k_1, k_2, \ldots, k_N)$

Graphs/Networks in a nutshell

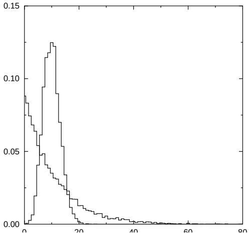
- size N; 'nodes': $i, j, k = 1 \ldots N$
- 'links': Connectivity matrix

$$c_{ij} = \begin{cases} 1 & i — j \\ 0 & otherwise \end{cases}$$

- degrees: $k_i = \sum_j c_{ij}$; $\mathbf{k} = (k_1, k_2, \ldots, k_N)$

- degree distribution

$$p(k) = N^{-1} \sum_i \delta_{k,k_i}$$



Graphs/Networks in a nutshell

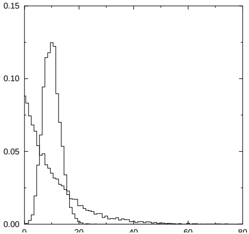
- size N; 'nodes': $i, j, k = 1...N$
- 'links': Connectivity matrix

$$c_{ij} = \begin{cases} 1 & i — j \\ 0 & otherwise \end{cases}$$

- degrees: $k_i = \sum_j c_{ij}$; $\quad \mathbf{k} = (k_1, k_2, \ldots, k_N)$

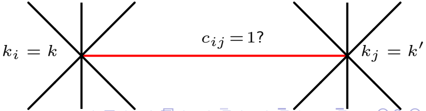
- degree distribution

$$p(k) = N^{-1} \sum_i \delta_{k,k_i}$$



- Degree correlation

$$W(k,k') = \frac{\sum_{ij} c_{ij} \delta_{k,k_i} \delta_{k',k_j}}{\sum_{ij} c_{ij}}$$



Graphs/Networks in a nutshell

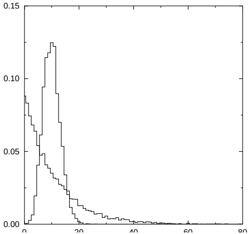
- size N; 'nodes': $i, j, k = 1 \ldots N$
- 'links': Connectivity matrix

$$c_{ij} = \begin{cases} 1 & i — j \\ 0 & otherwise \end{cases}$$

- degrees: $k_i = \sum_j c_{ij}$; $\quad \mathbf{k} = (k_1, k_2, \ldots, k_N)$

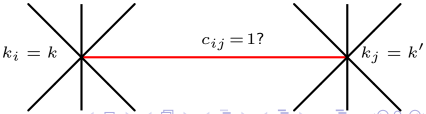
- degree distribution

$$p(k) = N^{-1} \sum_i \delta_{k,k_i}$$



- Degree correlation

$$W(k,k') = \frac{\sum_{ij} c_{ij} \delta_{k,k_i} \delta_{k',k_j}}{\sum_{ij} c_{ij}}$$



Macroscopic measures of topology

- $p(k)$, $W(k,k')$ macroscopic measures, independent on N

Macroscopic measures of topology

- $p(k)$, $W(k,k')$ macroscopic measures, independent on N
- Not independent:

$$W(k)=\sum_{k'} W(k,k')=\frac{kp(k)}{\bar{k}} \qquad \bar{k}=\sum_{k} kp(k)$$

Macroscopic measures of topology

- $p(k)$, $W(k,k')$ macroscopic measures, independent on N
- Not independent:

$$W(k) = \sum_{k'} W(k,k') = \frac{kp(k)}{\bar{k}} \qquad \bar{k} = \sum_{k} kp(k)$$

- No degree correlations ⇒ $W(k,k') = W(k)W(k')$

Macroscopic measures of topology

- $p(k)$, $W(k,k')$ macroscopic measures, independent on N
- Not independent:

$$W(k) = \sum_{k'} W(k,k') = \frac{kp(k)}{\bar{k}} \qquad \bar{k} = \sum_k kp(k)$$

- No degree correlations ⇒ $W(k,k') = W(k)W(k')$
- Define

$$\Pi(k,k') = W(k,k')/W(k)W(k')$$

Macroscopic measures of topology

- $p(k)$, $W(k,k')$ macroscopic measures, independent on N
- Not independent:

$$W(k)=\sum_{k'}W(k,k')=\frac{kp(k)}{\bar{k}} \qquad \bar{k}=\sum_{k}kp(k)$$

- No degree correlations $\Rightarrow$ $W(k,k')=W(k)W(k')$
- Define

$$\Pi(k,k')=W(k,k')/W(k)W(k')$$

$\Pi\neq 1$ signals presence of structure beyond degrees statistics

Tailored random graphs ensembles

- Random graph ensemble G: set of allowed graphs G and probability $P(\mathbf{c}) > 0 \ \forall \ \mathbf{c} \in G$

Tailored random graphs ensembles

- Random graph ensemble G: set of allowed graphs G and probability $P(\mathbf{c}) > 0 \ \forall \ \mathbf{c} \in G$
- Tailored G_L: given $\boldsymbol{\Omega}(\mathbf{c}) \equiv (\Omega_1(\mathbf{c}), \ldots, \Omega_L(\mathbf{c}))$, demand

Tailored random graphs ensembles

- Random graph ensemble G: set of allowed graphs G and probability $P(\mathbf{c}) > 0 \; \forall \; \mathbf{c} \in G$
- Tailored G_L: given $\mathbf{\Omega}(\mathbf{c}) \equiv (\Omega_1(\mathbf{c}), \ldots, \Omega_L(\mathbf{c}))$, demand
  - Hard: $\mathbf{\Omega}(\mathbf{c}) = \mathbf{\Omega} \; \forall \; \mathbf{c} \in G_L$

Tailored random graphs ensembles

- Random graph ensemble G: set of allowed graphs G and probability $P(\mathbf{c}) > 0 \ \forall \ \mathbf{c} \in G$
- Tailored G_L: given $\mathbf{\Omega}(\mathbf{c}) \equiv (\Omega_1(\mathbf{c}), \ldots, \Omega_L(\mathbf{c}))$, demand
  - Hard: $\mathbf{\Omega}(\mathbf{c}) = \mathbf{\Omega} \ \forall \ \mathbf{c} \in G_L$

$$P_L(\mathbf{c}|\mathbf{\Omega}) = \frac{1}{Z_L(\mathbf{\Omega})} \delta_{\mathbf{\Omega}(\mathbf{c}),\mathbf{\Omega}}, \qquad Z_L(\mathbf{\Omega}) = \sum_{\mathbf{c} \in G} \delta_{\mathbf{\Omega}(\mathbf{c}),\mathbf{\Omega}}$$

Tailored random graphs ensembles

- Random graph ensemble G: set of allowed graphs G and probability $P(\mathbf{c}) > 0 \ \forall \ \mathbf{c} \in G$
- Tailored G_L: given $\mathbf{\Omega}(\mathbf{c}) \equiv (\Omega_1(\mathbf{c}), \ldots, \Omega_L(\mathbf{c}))$, demand
  - Hard: $\mathbf{\Omega}(\mathbf{c}) = \mathbf{\Omega} \ \forall \ \mathbf{c} \in G_L$

$$P_L(\mathbf{c}|\mathbf{\Omega}) = \frac{1}{Z_L(\mathbf{\Omega})}\delta_{\mathbf{\Omega}(\mathbf{c}),\mathbf{\Omega}}, \qquad Z_L(\mathbf{\Omega}) = \sum_{\mathbf{c}\in G}\delta_{\mathbf{\Omega}(\mathbf{c}),\mathbf{\Omega}}$$

  - Soft: $\langle\Omega_\mu(\mathbf{c})\rangle = \Omega_\mu, \ \mu = 1, \ldots, L$

Tailored random graphs ensembles

- Random graph ensemble G: set of allowed graphs G and probability $P(\mathbf{c}) > 0 \ \forall \ \mathbf{c} \in G$
- Tailored G_L: given $\mathbf{\Omega}(\mathbf{c}) \equiv (\Omega_1(\mathbf{c}), \ldots, \Omega_L(\mathbf{c}))$, demand
  - Hard: $\mathbf{\Omega}(\mathbf{c}) = \mathbf{\Omega} \ \forall \ \mathbf{c} \in G_L$

$$P_L(\mathbf{c}|\mathbf{\Omega}) = \frac{1}{Z_L(\mathbf{\Omega})} \delta_{\mathbf{\Omega}(\mathbf{c}),\mathbf{\Omega}}, \qquad Z_L(\mathbf{\Omega}) = \sum_{\mathbf{c} \in G} \delta_{\mathbf{\Omega}(\mathbf{c}),\mathbf{\Omega}}$$

  - Soft: $\langle \Omega_\mu(\mathbf{c}) \rangle = \Omega_\mu$, $\mu = 1, \ldots, L$

$$P_L(\mathbf{c}|\mathbf{\Omega}) = \frac{1}{Z_L(\mathbf{\Omega})} e^{\sum_\mu \omega_\mu(\mathbf{\Omega}) \Omega_\mu(\mathbf{c})}, \qquad Z_L(\mathbf{\Omega}) = \sum_{\mathbf{c} \in G} e^{\sum_\mu \omega_\mu(\mathbf{\Omega}) \Omega_\mu(\mathbf{c})}$$

with $\omega_\mu(\mathbf{\Omega})$ solved from $\sum_{\mathbf{c} \in G} P_L(\mathbf{c}|\mathbf{\Omega}) \Omega_\mu(\mathbf{c}) = \Omega_\mu$

Tailored random graphs ensembles

- Random graph ensemble G: set of allowed graphs G and probability $P(\mathbf{c}) > 0\ \forall\ \mathbf{c} \in G$
- Tailored G_L: given $\mathbf{\Omega}(\mathbf{c}) \equiv (\Omega_1(\mathbf{c}), \ldots, \Omega_L(\mathbf{c}))$, demand
  - Hard: $\mathbf{\Omega}(\mathbf{c}) = \mathbf{\Omega}\ \forall\ \mathbf{c} \in G_L$

$$P_L(\mathbf{c}|\mathbf{\Omega}) = \frac{1}{Z_L(\mathbf{\Omega})}\delta_{\mathbf{\Omega}(\mathbf{c}),\mathbf{\Omega}}, \qquad Z_L(\mathbf{\Omega}) = \sum_{\mathbf{c}\in G}\delta_{\mathbf{\Omega}(\mathbf{c}),\mathbf{\Omega}}$$

  - Soft: $\langle \Omega_\mu(\mathbf{c})\rangle = \Omega_\mu$, $\mu = 1, \ldots, L$

$$P_L(\mathbf{c}|\mathbf{\Omega}) = \frac{1}{Z_L(\mathbf{\Omega})}e^{\sum_\mu \omega_\mu(\mathbf{\Omega})\Omega_\mu(\mathbf{c})}, \qquad Z_L(\mathbf{\Omega}) = \sum_{\mathbf{c}\in G}e^{\sum_\mu \omega_\mu(\mathbf{\Omega})\Omega_\mu(\mathbf{c})}$$

with $\omega_\mu(\mathbf{\Omega})$ solved from $\sum_{\mathbf{c}\in G} P_L(\mathbf{c}|\mathbf{\Omega})\Omega_\mu(\mathbf{c}) = \Omega_\mu$

Numerical sampling of $\omega_\mu(\mathbf{\Omega})$ hard for sophisticated $\mathbf{\Omega}$, but analytical progress feasible for suitable choices and N large!

Ensembles as null models or proxies

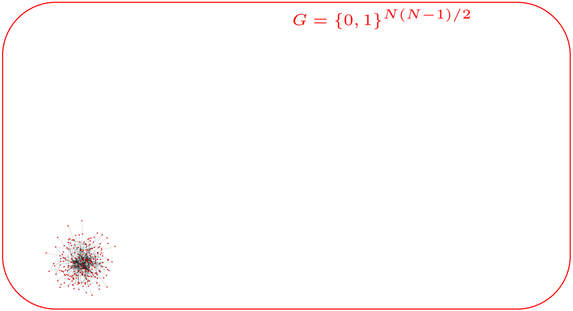
- Can approximate any $\mathbf{c}^\star$ by G_L with $\mathbf{\Omega} = \mathbf{\Omega}(\mathbf{c}^\star)$

Ensembles as null models or proxies

- Can approximate any $\mathbf{c}^{\star}$ by G_L with $\boldsymbol{\Omega} = \boldsymbol{\Omega}(\mathbf{c}^{\star})$
- Increasingly detailed measurements $\boldsymbol{\Omega} = (\bar{k}, p, \Pi, \ldots)$

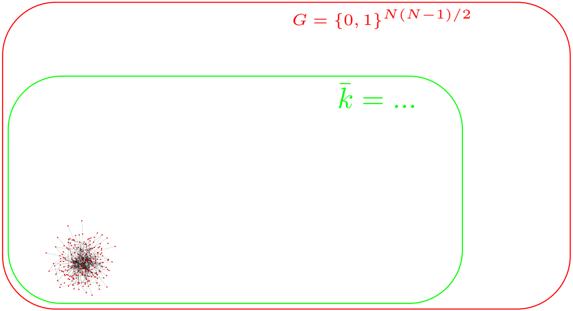
Ensembles as null models or proxies

- Can approximate any $\mathbf{c}^\star$ by G_L with $\mathbf{\Omega} = \mathbf{\Omega}(\mathbf{c}^\star)$
- Increasingly detailed measurements $\mathbf{\Omega} = (\bar{k}, p, \Pi, \ldots)$



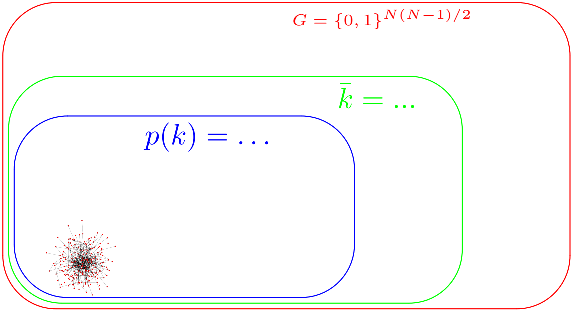
Ensembles as null models or proxies

- Can approximate any $\mathbf{c}^\star$ by G_L with $\mathbf{\Omega} = \mathbf{\Omega}(\mathbf{c}^\star)$
- Increasingly detailed measurements $\mathbf{\Omega} = (\bar{k}, p, \Pi, \ldots)$



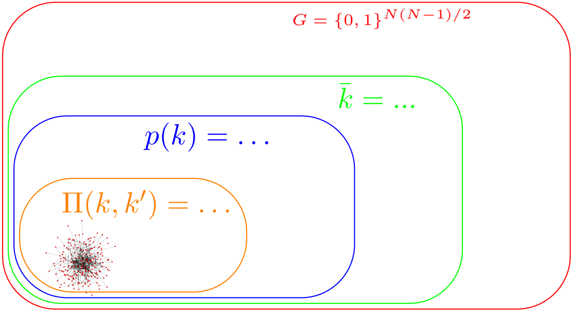
Ensembles as null models or proxies

- Can approximate any $\mathbf{c}^\star$ by G_L with $\mathbf{\Omega} = \mathbf{\Omega}(\mathbf{c}^\star)$
- Increasingly detailed measurements $\mathbf{\Omega} = (\bar{k}, p, \Pi, \ldots)$



Ensembles as null models or proxies

- Can approximate any $\mathbf{c}^\star$ by G_L with $\boldsymbol{\Omega} = \boldsymbol{\Omega}(\mathbf{c}^\star)$
- Increasingly detailed measurements $\boldsymbol{\Omega} = (\bar{k}, p, \Pi, \ldots)$



Hierarchy of ensembles

(i) prescribe only $\langle k\rangle$ Erdös-Renyi graphs

$$P(\mathbf{c}|\bar{k}) = \prod_{i<j}\left[\frac{\bar{k}}{N}\delta_{c_{ij},1} + \left(1-\frac{\bar{k}}{N}\right)\delta_{c_{ij},0}\right] \qquad \text{(soft)}$$

Hierarchy of ensembles

(i) prescribe only $\langle k \rangle$ Erdös-Renyi graphs

$$P(\mathbf{c}|\bar{k}) = \prod_{i<j} \left[\frac{\bar{k}}{N} \delta_{c_{ij},1} + \left(1 - \frac{\bar{k}}{N} \right) \delta_{c_{ij},0} \right] \qquad \text{(soft)}$$

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

$$P(\mathbf{c}|\mathbf{k}) = \frac{\delta_{\mathbf{k},\mathbf{k}(\mathbf{c})}}{Z(\mathbf{k})} \qquad \text{(hard)}$$

Hierarchy of ensembles

(i) prescribe only $\langle k \rangle$ Erdös-Renyi graphs

$$P(\mathbf{c}|\bar{k}) = \prod_{i<j} \left[\frac{\bar{k}}{N} \delta_{c_{ij},1} + \left(1 - \frac{\bar{k}}{N}\right) \delta_{c_{ij},0} \right] \qquad \text{(soft)}$$

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

$$P(\mathbf{c}|\mathbf{k}) = \frac{\delta_{\mathbf{k},\mathbf{k}(\mathbf{c})}}{Z(\mathbf{k})} \qquad \text{(hard)}$$

(iii) prescribe $(k_1, \ldots, k_N)$ and $\Pi(k, k')$?

Hierarchy of ensembles

(i) prescribe only $\langle k \rangle$ Erdös-Renyi graphs

$$P(\mathbf{c}|\bar{k}) = \prod_{i<j} \left[\frac{\bar{k}}{N} \delta_{c_{ij},1} + \left(1 - \frac{\bar{k}}{N}\right) \delta_{c_{ij},0} \right] \qquad \text{(soft)}$$

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

$$P(\mathbf{c}|\mathbf{k}) = \frac{\delta_{\mathbf{k},\mathbf{k}(\mathbf{c})}}{Z(\mathbf{k})} \qquad \text{(hard)}$$

(iii) prescribe $(k_1, \ldots, k_N)$ and $\Pi(k, k')$?

$$P(\mathbf{c}|\mathbf{k}, Q) = \frac{\delta_{\mathbf{k},\mathbf{k}(\mathbf{c})}}{Z(\mathbf{k}, Q)} \prod_{i<j} \left[\frac{\langle k \rangle}{N} Q(k_i, k_j) \delta_{c_{ij},1} + \left(1 - \frac{\langle k \rangle}{N} Q(k_i, k_j)\right) \delta_{c_{ij},0} \right]$$

Hierarchy of ensembles

(i) prescribe only $\langle k\rangle$ Erdös-Renyi graphs

$$P(\mathbf{c}|\bar{k}) = \prod_{i<j}\left[\frac{\bar{k}}{N}\delta_{c_{ij},1} + \left(1-\frac{\bar{k}}{N}\right)\delta_{c_{ij},0}\right] \qquad \text{(soft)}$$

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

$$P(\mathbf{c}|\mathbf{k}) = \frac{\delta_{\mathbf{k},\mathbf{k}(\mathbf{c})}}{Z(\mathbf{k})} \qquad \text{(hard)}$$

(iii) prescribe $(k_1,\ldots,k_N)$ and $\Pi(k,k')$?

$$P(\mathbf{c}|\mathbf{k},Q) = \frac{\delta_{\mathbf{k},\mathbf{k}(\mathbf{c})}}{Z(\mathbf{k},Q)}\prod_{i<j}\left[\frac{\langle k\rangle}{N}Q(k_i,k_j)\delta_{c_{ij},1} + \left(1-\frac{\langle k\rangle}{N}Q(k_i,k_j)\right)\delta_{c_{ij},0}\right]$$

Right choice:
$Q(k,k') = \Pi(k,k')kk'/\langle k\rangle^2$

Hierarchy of ensembles

(i) prescribe only $\langle k \rangle$ Erdös-Renyi graphs

$$P(\mathbf{c}|\bar{k}) = \prod_{i<j} \left[\frac{\bar{k}}{N} \delta_{c_{ij},1} + \left(1 - \frac{\bar{k}}{N}\right) \delta_{c_{ij},0} \right] \qquad \text{(soft)}$$

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

$$P(\mathbf{c}|\mathbf{k}) = \frac{\delta_{\mathbf{k},\mathbf{k}(\mathbf{c})}}{Z(\mathbf{k})} \qquad \text{(hard)}$$

(iii) prescribe $(k_1, \ldots, k_N)$ and $\Pi(k, k')$?

$$P(\mathbf{c}|\mathbf{k}, Q) = \frac{\delta_{\mathbf{k},\mathbf{k}(\mathbf{c})}}{Z(\mathbf{k}, Q)} \prod_{i<j} \left[\frac{\langle k \rangle}{N} Q(k_i, k_j) \delta_{c_{ij},1} + \left(1 - \frac{\langle k \rangle}{N} Q(k_i, k_j)\right) \delta_{c_{ij},0} \right]$$

Right choice:
$Q(k, k') = \Pi(k, k') k k' / \langle k \rangle^2 = W(k, k') / p(k) p(k')$

Modelling biological networks

- Biological network $\mathbf{c}$

Modelling biological networks

- Biological network $\mathbf{c} \Rightarrow$ measure $p(k), W(k, k')$

Modelling biological networks

- Biological network $\mathbf{c}$ ⇒ measure $p(k), W(k,k')$
- consider all graphs with same p, W

Modelling biological networks

- Biological network $\mathbf{c}$ ⇒ measure $p(k), W(k,k')$
- consider all graphs with same p, W

$$P(\mathbf{c}|p,W) = \sum_{\mathbf{k}} P(\mathbf{c}|\mathbf{k},W)\prod_i p(k_i)$$

$$P(\mathbf{c}|\mathbf{k},W) = \frac{1}{Z_N(\mathbf{k},W)}\Big[\prod_i \delta_{k_i,k_i(\mathbf{c})}\Big]$$

$$\times\prod_{i<j}\Big[\frac{\overline{k}}{N}\frac{W(k_i,k_j)}{p(k_i)p(k_j)}\delta_{c_{ij},1} + \Big(1-\frac{\overline{k}}{N}\frac{W(k_i,k_j)}{p(k_i)p(k_j)}\Big)\delta_{c_{ij},0}\Big]$$

[A Annibale, ACC Coolen, LP Fernandes, F Fraternali J Kleinjung *J. Phys. A: Math. Theor. 42 485001* (2009)]

Distance between networks

Information theory

⇓

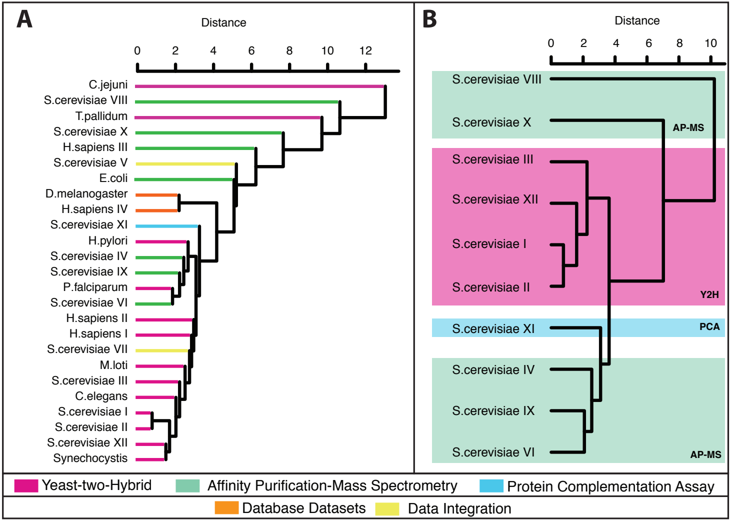
Distance between $\mathbf{c}_A$ and $\mathbf{c}_B$

$$D_{AB} = \frac{1}{2N}\sum_{\mathbf{c}} P(\mathbf{c}|p_A, W_A)\log\frac{P(\mathbf{c}|p_A, W_A)}{P(\mathbf{c}|p_B, W_B)}$$

$$+\frac{1}{2N}\sum_{\mathbf{c}} P(\mathbf{c}|p_B, W_B)\log\frac{P(\mathbf{c}|p_B, W_B)}{P(\mathbf{c}|p_A, W_A)}$$

$$= f(p_A, p_B, W_A, W_B)$$

Dendrograms



[LP Fernandes, A Annibale, J Kleinjung, ACC Coolen, F Fraternali *PLoS ONE 5(8): e12083* (2010)]

Accounting for biases

$$\begin{tikzcd}
\mathbf{c} \arrow[r, "\text{sampling}"] \arrow[d] & \mathbf{c}' \arrow[d] \\
p, W \arrow[r, "\text{relation?}"] & p', W'
\end{tikzcd}$$

Accounting for biases

$$\begin{tikzcd}
\mathbf{c} \arrow[r, "\text{sampling}"] \arrow[d] & \mathbf{c}' \arrow[d] \\
p, W \arrow[r, "\text{relation?}"] & p', W'
\end{tikzcd}$$

So far:

- only p' was studied and
- only for random node sampling

Accounting for biases

$$\begin{tikzcd}
\mathbf{c} \arrow[r, "\text{sampling}"] \arrow[d] & \mathbf{c}' \arrow[d] & \text{Microscopic} \\
p, W \arrow[r, "\text{relation?}"] & p', W' & \text{Macroscopic}
\end{tikzcd}$$

Outline

1. Motivation
2. Quantifying biases
    - Tailored random graph ensembles
    - Sampling protocols
    - Results
3. Inferring the true network from imperfect data
    - Bayesian analysis

(Connectivity-dependent) Sampling protocols

- node undersampling: $c'_{ij} = \sigma_i \sigma_j c_{ij}$

$$\sigma_i = \begin{cases} 1 & \text{node i detected} \\ 0 & \text{otherwise} \end{cases}$$

(Connectivity-dependent) Sampling protocols

- node undersampling: $c'_{ij} = \sigma_i\sigma_j c_{ij}$

$$\sigma_i = \begin{cases} 1 & \text{node i detected} \\ 0 & \text{otherwise} \end{cases}$$

- bond undersampling: $c'_{ij} = \tau_{ij} c_{ij}$

$$\tau_{ij} = \begin{cases} 1 & \text{bond } i-j \text{ detected} \\ 0 & \text{otherwise} \end{cases}$$

(Connectivity-dependent) Sampling protocols

- node undersampling: $c'_{ij} = \sigma_i \sigma_j c_{ij}$

$$\sigma_i = \begin{cases} 1 & \text{node i detected} \\ 0 & \text{otherwise} \end{cases}$$

- bond undersampling: $c'_{ij} = \tau_{ij} c_{ij}$

$$\tau_{ij} = \begin{cases} 1 & \text{bond } i - j \text{ detected} \\ 0 & \text{otherwise} \end{cases}$$

- bond oversampling: $c'_{ij} = c_{ij} + (1 - c_{ij})\lambda_{ij}$

$$\lambda_{ij} = \begin{cases} 1 & \text{bond } i - j \text{ created} \\ 0 & \text{otherwise} \end{cases}$$

(Connectivity-dependent) Sampling protocols

- node undersampling: $c'_{ij} = \sigma_i \sigma_j c_{ij}$

$$\sigma_i = \begin{cases} 1 & \text{node i detected} \\ 0 & \text{otherwise} \end{cases}$$

- bond undersampling: $c'_{ij} = \tau_{ij} c_{ij}$

$$\tau_{ij} = \begin{cases} 1 & \text{bond } i-j \text{ detected} \\ 0 & \text{otherwise} \end{cases}$$

- bond oversampling: $c'_{ij} = c_{ij} + (1 - c_{ij})\lambda_{ij}$

$$\lambda_{ij} = \begin{cases} 1 & \text{bond } i-j \text{ created} \\ 0 & \text{otherwise} \end{cases}$$

$$\Downarrow \qquad \text{in combination}$$

$$c'_{ij} = \sigma_i \sigma_j [\tau_{ij} c_{ij} + (1 - c_{ij})\lambda_{ij}]$$

(Connectivity-dependent) Sampling protocols

- node undersampling: $c'_{ij} = \sigma_i \sigma_j c_{ij}$

$$\sigma_i = \begin{cases} 1 & \text{node i detected} \qquad\qquad x \\ 0 & \text{otherwise} \qquad\qquad\quad 1-x \end{cases}$$

- bond undersampling: $c'_{ij} = \tau_{ij} c_{ij}$

$$\tau_{ij} = \begin{cases} 1 & \text{bond } i-j \text{ detected} \\ 0 & \text{otherwise} \end{cases}$$

- bond oversampling: $c'_{ij} = c_{ij} + (1-c_{ij})\lambda_{ij}$

$$\lambda_{ij} = \begin{cases} 1 & \text{bond } i-j \text{ created} \\ 0 & \text{otherwise} \end{cases}$$

$$\Downarrow \qquad \text{in combination}$$

$$c'_{ij} = \sigma_i \sigma_j [\tau_{ij} c_{ij} + (1-c_{ij})\lambda_{ij}]$$

(Connectivity-dependent) Sampling protocols

- node undersampling: $c'_{ij} = \sigma_i \sigma_j c_{ij}$

$$\sigma_i = \begin{cases} 1 & \text{node i detected} \qquad\qquad x \\ 0 & \text{otherwise} \qquad\qquad 1-x \end{cases}$$

- bond undersampling: $c'_{ij} = \tau_{ij} c_{ij}$

$$\tau_{ij} = \begin{cases} 1 & \text{bond } i-j \text{ detected} \qquad\qquad y \\ 0 & \text{otherwise} \qquad\qquad 1-y \end{cases}$$

- bond oversampling: $c'_{ij} = c_{ij} + (1-c_{ij})\lambda_{ij}$

$$\lambda_{ij} = \begin{cases} 1 & \text{bond } i-j \text{ created} \\ 0 & \text{otherwise} \end{cases}$$

$$\Downarrow \qquad \text{in combination}$$

$$c'_{ij} = \sigma_i \sigma_j [\tau_{ij} c_{ij} + (1-c_{ij})\lambda_{ij}]$$

(Connectivity-dependent) Sampling protocols

- node undersampling: $c'_{ij} = \sigma_i\sigma_j c_{ij}$

$$\sigma_i = \begin{cases} 1 & \text{node i detected} \qquad x \\ 0 & \text{otherwise} \qquad 1-x \end{cases}$$

- bond undersampling: $c'_{ij} = \tau_{ij} c_{ij}$

$$\tau_{ij} = \begin{cases} 1 & \text{bond } i-j \text{ detected} \qquad y \\ 0 & \text{otherwise} \qquad 1-y \end{cases}$$

- bond oversampling: $c'_{ij} = c_{ij} + (1-c_{ij})\lambda_{ij}$

$$\lambda_{ij} = \begin{cases} 1 & \text{bond } i-j \text{ created} \qquad N^{-1}z \\ 0 & \text{otherwise} \qquad 1-N^{-1}z \end{cases}$$

$\Downarrow$ in combination

$$c'_{ij} = \sigma_i\sigma_j[\tau_{ij}c_{ij} + (1-c_{ij})\lambda_{ij}]$$

(Connectivity-dependent) Sampling protocols

- node undersampling: $c'_{ij} = \sigma_i \sigma_j c_{ij}$

$$\sigma_i = \begin{cases} 1 & \text{node i detected} \qquad x(k_i) \\ 0 & \text{otherwise} \qquad 1 - x(k_i) \end{cases}$$

- bond undersampling: $c'_{ij} = \tau_{ij} c_{ij}$

$$\tau_{ij} = \begin{cases} 1 & \text{bond } i-j \text{ detected} \qquad y(k_i, k_j) \\ 0 & \text{otherwise} \qquad 1 - y(k_i, k_j) \end{cases}$$

- bond oversampling: $c'_{ij} = c_{ij} + (1 - c_{ij})\lambda_{ij}$

$$\lambda_{ij} = \begin{cases} 1 & \text{bond } i-j \text{ created} \qquad N^{-1} z(k_i, k_j) \\ 0 & \text{otherwise} \qquad 1 - N^{-1} z(k_i, k_j) \end{cases}$$

$$\Downarrow \qquad \text{in combination}$$

$$c'_{ij} = \sigma_i \sigma_j [\tau_{ij} c_{ij} + (1 - c_{ij})\lambda_{ij}]$$

Macroscopic features

$$\mathbf{c} \xrightarrow{\boldsymbol{\sigma},\,\boldsymbol{\tau},\,\boldsymbol{\lambda}} \mathbf{c}'$$

Macroscopic features

$$\begin{tikzcd} \mathbf{c} \arrow[r, "{\boldsymbol{\sigma}, \boldsymbol{\tau}, \boldsymbol{\lambda}}"] \arrow[d] & \mathbf{c}' \\ p_0, W_0 & \end{tikzcd}$$

Macroscopic features

$$\begin{tikzcd}
\mathbf{c} \arrow[r, "{\boldsymbol{\sigma}, \boldsymbol{\tau}, \boldsymbol{\lambda}}"] \arrow[d] & \mathbf{c}' \arrow[d] \\
p_0, W_0 & p, W
\end{tikzcd}$$

$$p(k|\mathbf{c}') = \frac{\sum_i \sigma_i \delta_{k,\sum_j c_{ij'}}}{\sum_i \sigma_i}$$

$$W(k, k'|\mathbf{c}') = \frac{\sum_{ij} c'_{ij} \delta_{k,\sum_\ell c'_{i\ell}} \delta_{k',\sum_\ell c'_{j\ell}}}{\sum_{ij} c'_{ij}}$$

Macroscopic features

$$\begin{tikzcd}
\mathbf{c} \arrow[r, "\boldsymbol{\sigma},\boldsymbol{\tau},\boldsymbol{\lambda}"] \arrow[d] & \mathbf{c}' \arrow[d, "\langle\cdots\rangle_{\boldsymbol{\sigma},\boldsymbol{\tau},\boldsymbol{\lambda}}"] \\
p_0, W_0 & p, W
\end{tikzcd}$$

$$p(k|\mathbf{c}') = \frac{\sum_i \sigma_i \delta_{k,\sum_j c_{ij'}}}{\sum_i \sigma_i}$$

$$W(k,k'|\mathbf{c}') = \frac{\sum_{ij} c'_{ij} \delta_{k,\sum_\ell c'_{i\ell}} \delta_{k',\sum_\ell c'_{j\ell}}}{\sum_{ij} c'_{ij}}$$

Macroscopic features

$$\begin{tikzcd}
\mathbf{c} \arrow[r, "\boldsymbol{\sigma}, \boldsymbol{\tau}, \boldsymbol{\lambda}"] \arrow[d] & \mathbf{c}' \arrow[d, "\langle \cdots \rangle_{\boldsymbol{\sigma},\boldsymbol{\tau},\boldsymbol{\lambda}}"] \\
p_0, W_0 & {}
\end{tikzcd}$$

$$p(k|\mathbf{c}') = \left\langle \frac{\sum_i \sigma_i \delta_{k,\sum_j c_{ij'}}}{\sum_i \sigma_i} \right\rangle_{\boldsymbol{\sigma},\boldsymbol{\tau},\boldsymbol{\lambda}}$$

$$W(k,k'|\mathbf{c}') = \left\langle \frac{\sum_{ij} c'_{ij} \delta_{k,\sum_\ell c'_{i\ell}} \delta_{k',\sum_\ell c'_{j\ell}}}{\sum_{ij} c'_{ij}} \right\rangle_{\boldsymbol{\sigma},\boldsymbol{\tau},\boldsymbol{\lambda}}$$

Macroscopic features

$$\begin{tikzcd}
\mathbf{c} \arrow[r, "{\boldsymbol{\sigma}, \boldsymbol{\tau}, \boldsymbol{\lambda}}"] \arrow[d] & \mathbf{c}' \arrow[d, "{\langle \cdots \rangle_{\boldsymbol{\sigma},\boldsymbol{\tau},\boldsymbol{\lambda}}}"] \\
p_0, W_0 & ?
\end{tikzcd}$$

$$p(k|\mathbf{c}') = \left\langle \frac{\sum_i \sigma_i \delta_{k,\sum_j c_{ij'}}}{\sum_i \sigma_i} \right\rangle_{\boldsymbol{\sigma},\boldsymbol{\tau},\boldsymbol{\lambda}}$$

$$W(k, k'|\mathbf{c}') = \left\langle \frac{\sum_{ij} c'_{ij} \delta_{k,\sum_\ell c'_{i\ell}} \delta_{k',\sum_\ell c'_{j\ell}}}{\sum_{ij} c'_{ij}} \right\rangle_{\boldsymbol{\sigma},\boldsymbol{\tau},\boldsymbol{\lambda}}$$

Macroscopic features

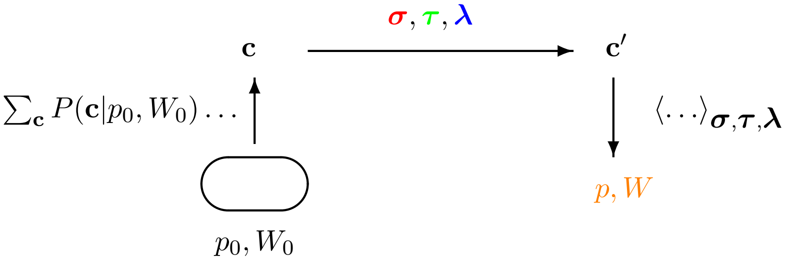
$$\begin{tikzcd}
\mathbf{c} \arrow[r, "\boldsymbol{\sigma},\boldsymbol{\tau},\boldsymbol{\lambda}"] & \mathbf{c}' \arrow[d, "\langle\cdots\rangle_{\boldsymbol{\sigma},\boldsymbol{\tau},\boldsymbol{\lambda}}"] \\
\boxed{} & ?
\end{tikzcd}$$

p_0, W_0

$$p(k|\mathbf{c}') = \left\langle \frac{\sum_i \sigma_i \delta_{k,\sum_j c_{ij'}}}{\sum_i \sigma_i} \right\rangle_{\boldsymbol{\sigma},\boldsymbol{\tau},\boldsymbol{\lambda}}$$

$$W(k,k'|\mathbf{c}') = \left\langle \frac{\sum_{ij} c'_{ij} \delta_{k,\sum_\ell c'_{i\ell}} \delta_{k',\sum_\ell c'_{j\ell}}}{\sum_{ij} c'_{ij}} \right\rangle_{\boldsymbol{\sigma},\boldsymbol{\tau},\boldsymbol{\lambda}}$$

Macroscopic features



$\boldsymbol{\sigma}, \boldsymbol{\tau}, \boldsymbol{\lambda}$

$\mathbf{c}$

$\mathbf{c}'$

$\sum_{\mathbf{c}} P(\mathbf{c}|p_0, W_0) \ldots$

$\langle \cdots \rangle_{\boldsymbol{\sigma},\boldsymbol{\tau},\boldsymbol{\lambda}}$

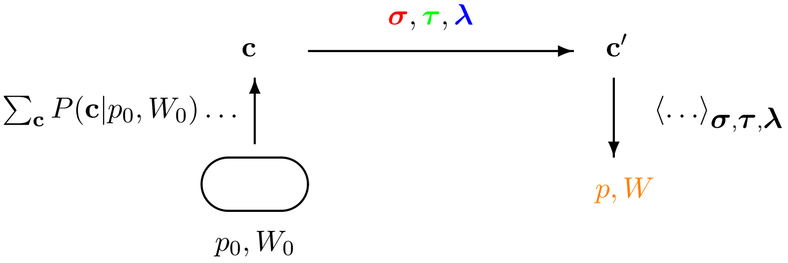
p, W

p_0, W_0

$$p(k|x,y,z) = \lim_{N\to\infty} \sum_{\mathbf{c}} P(\mathbf{c}|p_0, W_0) \left\langle \frac{\sum_i \sigma_i \delta_{k,\sum_j c_{ij'}}}{\sum_i \sigma_i} \right\rangle_{\boldsymbol{\sigma},\boldsymbol{\tau},\boldsymbol{\lambda}}$$

$$W(k,k'|x,y,z) = \lim_{N\to\infty} \sum_{\mathbf{c}} P(\mathbf{c}|p_0, W_0) \left\langle \frac{\sum_{ij} c'_{ij} \delta_{k,\sum_\ell c'_{i\ell}} \delta_{k',\sum_\ell c'_{j\ell}}}{\sum_{ij} c'_{ij}} \right\rangle_{\boldsymbol{\sigma},\boldsymbol{\tau},\boldsymbol{\lambda}}$$

Macroscopic features



$\boldsymbol{\sigma}, \boldsymbol{\tau}, \boldsymbol{\lambda}$

$\mathbf{c}$

$\mathbf{c}'$

$\sum_{\mathbf{c}} P(\mathbf{c}|p_0, W_0) \ldots$

$\langle \cdots \rangle_{\boldsymbol{\sigma},\boldsymbol{\tau},\boldsymbol{\lambda}}$

p, W

p_0, W_0

$$p(k|x,y,z) = \lim_{N\to\infty} \sum_{\mathbf{c}} P(\mathbf{c}|p_0, W_0) \left\langle \frac{\sum_i \sigma_i \delta_{k,\sum_j c_{ij'}}}{\sum_i \sigma_i} \right\rangle_{\boldsymbol{\sigma},\boldsymbol{\tau},\boldsymbol{\lambda}}$$

$$W(k,k'|x,y,z) = \lim_{N\to\infty} \sum_{\mathbf{c}} P(\mathbf{c}|p_0, W_0) \left\langle \frac{\sum_{ij} c'_{ij} \delta_{k,\sum_\ell c'_{i\ell}} \delta_{k',\sum_\ell c'_{j\ell}}}{\sum_{ij} c'_{ij}} \right\rangle_{\boldsymbol{\sigma},\boldsymbol{\tau},\boldsymbol{\lambda}}$$

$\Rightarrow$ Statistical mechanics techniques $\Rightarrow$

Macroscopic features

$$\begin{tikzcd}
\mathbf{c} \arrow[rr, "\boldsymbol{\sigma},\boldsymbol{\tau},\boldsymbol{\lambda}"] & & \mathbf{c}' \arrow[d, "\langle\cdots\rangle_{\boldsymbol{\sigma},\boldsymbol{\tau},\boldsymbol{\lambda}}"] \\
\underset{p_0,W_0}{\boxed{}} \arrow[u, "\sum_{\mathbf{c}} P(\mathbf{c}|p_0,W_0)\ldots"'] \arrow[rr, "x,y,z"] & & p,W
\end{tikzcd}$$

$$p(k|x,y,z)=\lim_{N\to\infty}\sum_{\mathbf{c}}P(\mathbf{c}|p_0,W_0)\left\langle\frac{\sum_i \sigma_i\delta_{k,\sum_j c_{ij'}}}{\sum_i\sigma_i}\right\rangle_{\boldsymbol{\sigma},\boldsymbol{\tau},\boldsymbol{\lambda}}$$

$$W(k,k'|x,y,z)=\lim_{N\to\infty}\sum_{\mathbf{c}}P(\mathbf{c}|p_0,W_0)\left\langle\frac{\sum_{ij} c'_{ij}\delta_{k,\sum_\ell c'_{i\ell}}\delta_{k',\sum_\ell c'_{j\ell}}}{\sum_{ij}c'_{ij}}\right\rangle_{\boldsymbol{\sigma},\boldsymbol{\tau},\boldsymbol{\lambda}}$$

$\Rightarrow$ Statistical mechanics techniques $\Rightarrow$

Outline

1. Motivation
2. Quantifying biases
    - Tailored random graph ensembles
    - Sampling protocols
    - Results
3. Inferring the true network from imperfect data
    - Bayesian analysis

Results

$$p(k|x,y,z) = \frac{\sum_q x(q)p(q)\Big\{a(q)\mathcal{J}(k|q) + qb(q)\mathcal{L}(k|q)\Big\}}{k\sum_q p(q)x(q)}$$

$$W(k,k'|x,y,z) = \frac{\sum_{q,q'>0} x(q)x(q')\Big\{p(q)p(q')z(q,q')\mathcal{J}(k|q)\mathcal{J}(k'|q') + \overline{k}W(q,q')y(q,q')\mathcal{L}(k|q)\mathcal{L}(k'|q')\Big\}}{\bar{k}(x,y,z)\sum_q p(q)x(q)}$$

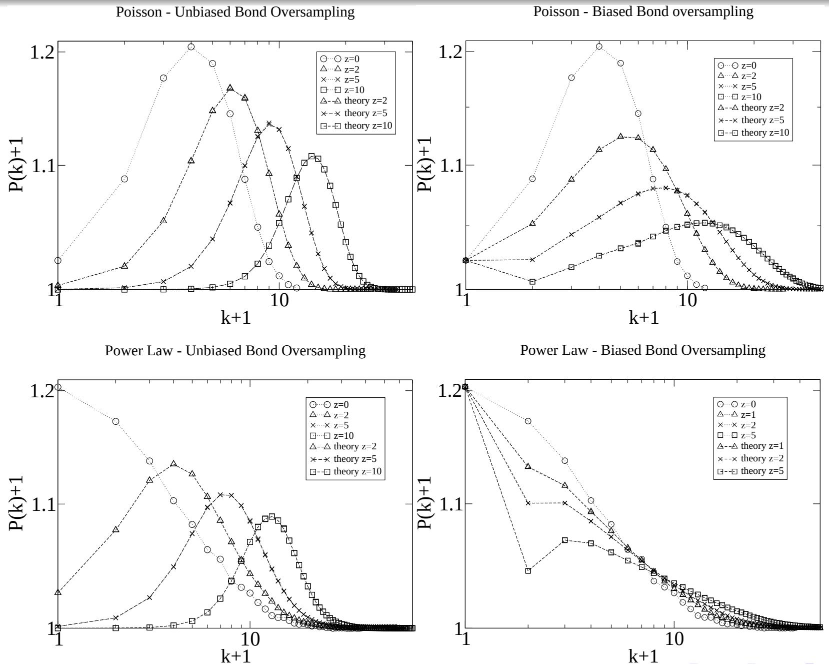
$$\bar{k}(x,y,z) = \sum_k k\,p(k|x,y,z) = \frac{\sum_q x(q)p(q)[a(q) + q\,b(q)]}{\sum_q p(q)x(q)}$$

with

$$\mathcal{J}(k|q) = e^{-a(q)} \sum_{n=0}^{\min\{k-1,q\}} \left(\begin{array}{c} q \\ n \end{array}\right) \frac{a^{k-1-n}(q)}{(k-1-n)!} b^n(q)(1-b(q))^{q-n}$$

$$\mathcal{L}(k|q) = e^{-a(q)} \sum_{n=0}^{\min\{k-1,q-1\}} \left(\begin{array}{c} q-1 \\ n \end{array}\right) \frac{a^{k-1-n}(q)}{(k-1-n)!} b^n(q)(1-b(q))^{q-1-n}$$

$$a(q) = \sum_{q'\geq 0} p(q')x(q')z(q,q'), \qquad b(q) = \frac{\overline{k}}{qp(q)} \sum_{q'\geq 0} x(q')y(q,q')W(q,q')$$



Poisson - Unbiased Bond Oversampling

Poisson - Biased Bond oversampling

Power Law - Unbiased Bond Oversampling

Power Law - Biased Bond Oversampling

Random sampling from Elegans: degree correlations

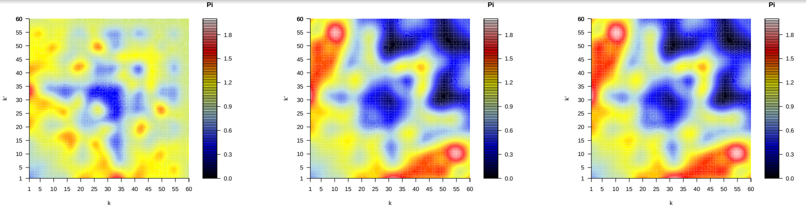


Figure: Random bond undersampling $x = 1, y = 0.9, z = 0,\ N = 3512,\ \bar{k} = 3.72$

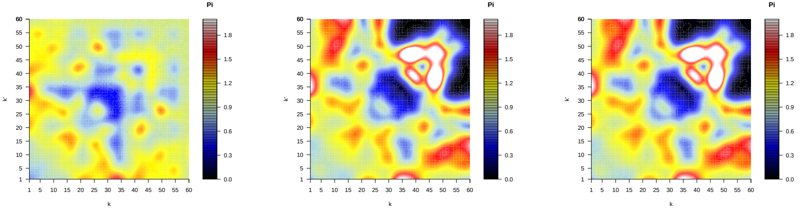


Figure: Random bond oversampling $x = 1, y = 1, z = 1,\ N = 3512,\ \bar{k} = 3.72$

[A Annibale, ACC Coolen *Interface Focus December 6, 2011 1:836-856*]

Outline

1. Motivation
2. Quantifying biases
   - Tailored random graph ensembles
   - Sampling protocols
   - Results
3. Inferring the true network from imperfect data
   - Bayesian analysis

Bayesian analysis

- $\ell = 1, ..., L$ species

Bayesian analysis

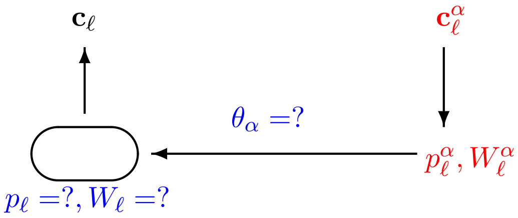
- $\ell = 1, ..., L$ species
- $\alpha = 1, ..., M$ experimental protocols, parameters $\theta_\alpha = \{x_\alpha, y_\alpha, z_\alpha\}$

Bayesian analysis

- $\ell = 1, ..., L$ species
- $\alpha = 1, ..., M$ experimental protocols, parameters $\theta_\alpha = \{x_\alpha, y_\alpha, z_\alpha\}$
- Observed networks $c_\ell^\alpha \Rightarrow p_\ell, W_\ell, \theta_\alpha$?

Bayesian analysis

- $\ell = 1, ..., L$ species
- $\alpha = 1, ..., M$ experimental protocols, parameters $\theta_\alpha = \{x_\alpha, y_\alpha, z_\alpha\}$
- Observed networks $c_\ell^\alpha \Rightarrow p_\ell, W_\ell, \theta_\alpha$?



$\mathbf{c}_\ell$

$\mathbf{c}_\ell^\alpha$

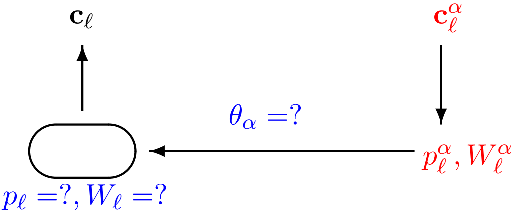
$\theta_\alpha =?$

$p_\ell^\alpha, W_\ell^\alpha$

$p_\ell =?, W_\ell =?$

Bayesian analysis

- $\ell = 1, ..., L$ species
- $\alpha = 1, ..., M$ experimental protocols, parameters $\theta_\alpha = \{x_\alpha, y_\alpha, z_\alpha\}$
- Observed networks $c_\ell^\alpha \Rightarrow p_\ell, W_\ell, \theta_\alpha$?



- Maximize $p(\{\theta_\alpha\}, \{p_\ell\}, \{W_\ell\} | \{\mathbf{c}_\ell^\alpha\})$ over $p_\ell, W_\ell, \theta_\alpha$

- Uniform choice for prior

$$p(p_\ell, W_\ell), \quad p(\theta_\alpha)$$

- Uniform choice for prior

$$p(p_\ell, W_\ell), \quad p(\theta_\alpha)$$

- Likelihood:

$$p(\mathbf{c}_\ell^\alpha|\theta_\alpha, p_\ell, W_\ell) = \sum_{\mathbf{c}_\ell} P(\mathbf{c}_\ell|W_\ell, p_\ell)\ p(\mathbf{c}_\ell^\alpha|\theta_\alpha, \mathbf{c}_\ell)$$

- Uniform choice for prior

$$p(p_\ell, W_\ell), \quad p(\theta_\alpha)$$

- Likelihood:

$$p(\mathbf{c}_\ell^\alpha|\theta_\alpha, p_\ell, W_\ell) = \sum_{\mathbf{c}_\ell} P(\mathbf{c}_\ell|W_\ell, p_\ell)\; p(\mathbf{c}_\ell^\alpha|\theta_\alpha, \mathbf{c}_\ell)$$

with

$p(\mathbf{c}_\ell^\alpha|\theta^\alpha, \mathbf{c}_\ell)$ determined from relation between $\mathbf{c}_\ell$ and $\mathbf{c}_\ell^\alpha$ (known!)

- Uniform choice for prior

$$p(p_\ell, W_\ell), \quad p(\theta_\alpha)$$

- Likelihood:

$$p(\mathbf{c}_\ell^\alpha|\theta_\alpha, p_\ell, W_\ell) = \sum_{\mathbf{c}_\ell} P(\mathbf{c}_\ell|W_\ell, p_\ell)\ p(\mathbf{c}_\ell^\alpha|\theta_\alpha, \mathbf{c}_\ell)$$

with

$p(\mathbf{c}_\ell^\alpha|\theta^\alpha, \mathbf{c}_\ell)$ determined from relation between $\mathbf{c}_\ell$ and $\mathbf{c}_\ell^\alpha$ (known!)

- Calculate $\langle p(\mathbf{c}_\ell^\alpha|\theta^\alpha, \mathbf{c}_\ell)\rangle$ via statistical mechanics
- Maximise posterior using Lagrange multipliers to handle constraints

$$\sum_k p_\ell(k) = 1 \qquad \sum_q W_\ell(k,q) = kp_\ell(k)/\bar{k}_\ell$$

- Uniform choice for prior

$$p(p_\ell, W_\ell), \quad p(\theta_\alpha)$$

- Likelihood:

$$p(\mathbf{c}_\ell^\alpha | \theta_\alpha, p_\ell, W_\ell) = \sum_{\mathbf{c}_\ell} P(\mathbf{c}_\ell | W_\ell, p_\ell)\ p(\mathbf{c}_\ell^\alpha | \theta_\alpha, \mathbf{c}_\ell)$$

with

$p(\mathbf{c}_\ell^\alpha | \theta^\alpha, \mathbf{c}_\ell)$ determined from relation between $\mathbf{c}_\ell$ and $\mathbf{c}_\ell^\alpha$ (known!)

- Calculate $\langle p(\mathbf{c}_\ell^\alpha | \theta^\alpha, \mathbf{c}_\ell) \rangle$ via statistical mechanics
- Maximise posterior using Lagrange multipliers to handle constraints

$$\sum_k p_\ell(k) = 1 \qquad \sum_q W_\ell(k,q) = k p_\ell(k)/\bar{k}_\ell$$

- get a set of equations for p_ℓ, W_ℓ and $x^\alpha, y^\alpha, z^\alpha$ in terms of the observed $p_\ell^\alpha, W_\ell^\alpha$

Conclusions

- Tailored random graphs ensemble can be used to model complex networks and quantify distances between them.
- Tailored graph ensembles can be used to quantify sampling effects on degree distributions and degree correlations for general sampling protocols (simulations match theory!)
- Underway: Bayesian inference of macroscopic features of biological networks and sampling parameters of different experiments given the observed networks
- Future: go all the way back to the original matrices

Aknowledgements

- ACC Coolen (maths)
- LP Fernandes, F Fraternali, J Kleinjung (bioinformatics)

ACC Coolen, F Fraternali, A Annibale, LP Fernandes, J Kleinjung
Modelling Biological Networks via Tailored Random Graphs
in *Handbook of Statistical Systems Biology*, MPH Stumpf, DJ Balding, M Girolami, Wiley, 2011

A Annibale, ACC Coolen, LP Fernandes, F Fraternali J Kleinjung
Tailored graph ensembles as proxies or null models for real networks I: tools for quantifying structure
J. Phys. A: Math. Theor. **42** *485001 (2009)*

ACC Coolen, A De Martino, A Annibale
Constrained Markovian dynamics of random graphs
J. Stat. Phys. **136** *(2009), 1035-1067*

LP Fernandes, A Annibale, J Kleinjung, ACC Coolen, F Fraternali
Protein networks reveal detection bias and species consistency when viewd through information-theoretic glasses
PLoS ONE 5(8): e12083 (2010)

A Annibale, ACC Coolen
What you see is not what you get: how sampling affects macroscopic features of biological networks
Interface Focus (2011) **1**, *836-856*

A. Annibale, ACC Coolen

Infering protein interactions networks from biased experimental data (in preparation).