Structure and dynamics of complex networks

March 10, 2020

Full statistical description

Full statistical description of assortativity

- $E_{k',k}$: number of links between nodes of degree k' and k.
- $e_{k',k} = \frac{E_{k',k}}{2L}$: the probability for finding a node with degree k' at one end and a node with degree k at the other end of a randomly selected link,
- $q_k = \sum_{k'} e_{k',k}$: prob. of degree k at one end of a link,
- For a neutral network we expect $e_{k'k} = q_{k'} \cdot q_k$, deviations from this value signify the presence of degree correlations!

Full statistical description

- q_k: the probability for finding a node with degree k at one end of a randomly selected link.
- How can we express q_k with the help of $p(k)$?

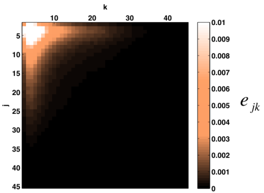
Full statistical description

- q_k: the probability for finding a node with degree k at one end of a randomly selected link.
- How can we express q_k with the help of $p(k)$?
  q_k is proportional to the total number of links hanging on nodes with degree k:
  $$q_k \sim kN_k$$

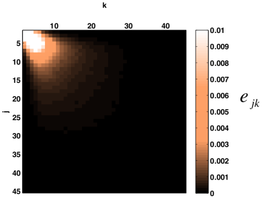
Full statistical description

- q_k: the probability for finding a node with degree k at one end of a randomly selected link.
- How can we express q_k with the help of $p(k)$?
  q_k is proportional to the total number of links hanging on nodes with degree k:
  $$q_k \sim kN_k$$
  To turn this into a normalized probability, we have to sum over all possibilities:
  $$q_k = \frac{kN_k}{\sum_{k'} k'N_{k'}} = \frac{kp(k)N}{\sum_{k'} k'p(k')N} = \frac{kp(k)}{\langle k \rangle}$$

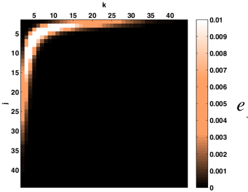
Averaging on 100 samples of SF networks with
$N = 10^4$, $\gamma = 2.5$, $\langle k \rangle = 3$:



Neutral



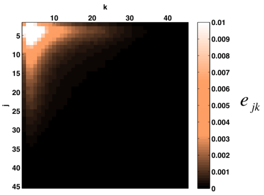
Assortative



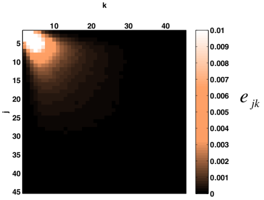
Disassortative

How would the distribution look like for a perfectly assortative/disassortative network?

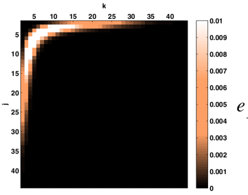
Averaging on 100 samples of SF networks with $N = 10^4$, $\gamma = 2.5$, $\langle k \rangle = 3$:



Neutral



Assortative

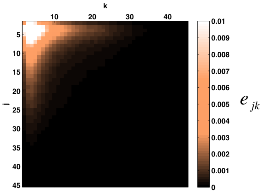


Disassortative

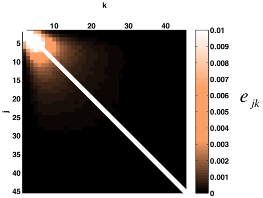
How would the distribution look like for a perfectly assortative/disassortative network?

Statistical description of assortativity

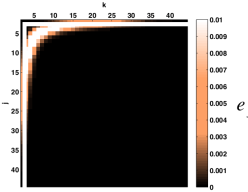
Averaging on 100 samples of SF networks with $N = 10^4$, $\gamma = 2.5$, $\langle k \rangle = 3$:



Neutral



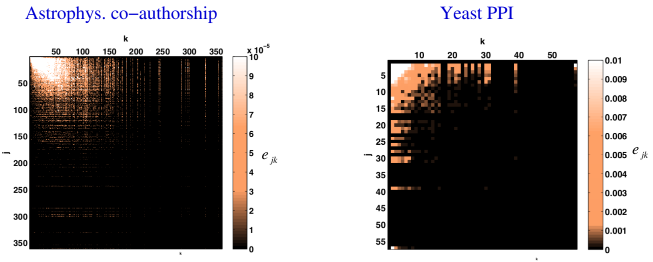
Assortative



Disassortative

How would the distribution look like for a perfectly assortative/disassortative network?

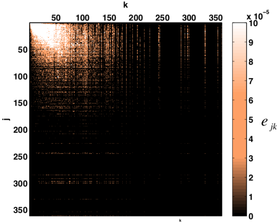
Results for real networks:



Statistical description of assortativity

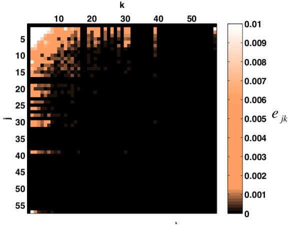
Results for real networks:

Astrophys. co–authorship



assortative

Yeast PPI



disassortative

Statistical description of assortativity

Problems with e_{kl}:

- difficult to prepare,
- difficult to evaluate.

ANND

How could we simplify the description of the degree correlations?

ANND

How could we simplify the description of the degree correlations?

→ **A**verage **N**earest **N**eighbors **D**egree!

ANND

ANND (or $k_{\rm nn}$) of individual nodes

- The ANND of node i:

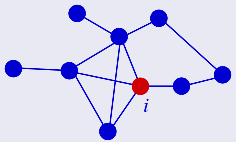
$$k_i^{\rm ANND} \equiv k_{{\rm nn},i} \equiv \left\langle k_j \right\rangle_{j \text{ linked to } i} = \frac{1}{k_i} \sum_{j \text{ linked to } i} k_j.$$

ANND

ANND (or $k_{\rm nn}$) of individual nodes

- The ANND of node i:

$$k_i^{\rm ANND} \equiv k_{{\rm nn},i} \equiv \left\langle k_j \right\rangle_{j \text{ linked to } i} = \frac{1}{k_i} \sum_{j \text{ linked to } i} k_j .$$



$$k_{{\rm nn},i} = \frac{2+5+4+3}{4} = 3.5$$

ANND

ANND of the network

- Once we calculated the ANND for every node, we can calculate further averages, e.g., what is the ANND for nodes with degree k?

→ The ANND of the whole network:

$$k_{\mathrm{nn}}(k) \equiv \langle k_{\mathrm{nn},i} \rangle_{k_i = k} .$$

- In terms of $P(k' \mid k)$ and $e_{k'k}$:

$$k_{\mathrm{nn}}(k) = \sum_{k'} k' P(k' \mid k), \qquad P(k' \mid k) = \frac{E_{k'k}}{\sum_{k'} E_{k'k}} = \frac{e_{k'k}}{\sum_{k'} e_{k'k}}$$

$$\rightarrow \quad k_{\mathrm{nn}}(k) = \frac{\sum_{k'} k' e_{k'k}}{\sum_{k'} e_{k'k}} = \frac{\sum_{k'} k' e_{k'k}}{q_k}$$

ANND

ANND of the network

- Once we calculated the ANND for every node, we can calculate further averages, e.g., what is the ANND for nodes with degree k?

→ The ANND of the whole network:

$$k_{\rm nn}(k) \equiv \left\langle k_{{\rm nn},i} \right\rangle_{k_i=k}.$$

- In terms of $P(k' \mid k)$ and $e_{k'k}$:

$$k_{\rm nn}(k) = \sum_{k'} k' P(k' \mid k), \qquad P(k' \mid k) = \frac{E_{k'k}}{\sum_{k'} E_{k'k}} = \frac{e_{k'k}}{\sum_{k'} e_{k'k}}$$

$$\rightarrow \quad k_{\rm nn}(k) = \frac{\sum_{k'} k' e_{k'k}}{\sum_{k'} e_{k'k}} = \frac{\sum_{k'} k' e_{k'k}}{q_k}$$

ANND

ANND of the network

- Once we calculated the ANND for every node, we can calculate further averages, e.g., what is the ANND for nodes with degree k?

→ The ANND of the whole network:

$$k_{\rm nn}(k) \equiv \left\langle k_{{\rm nn},i} \right\rangle_{k_i = k}.$$

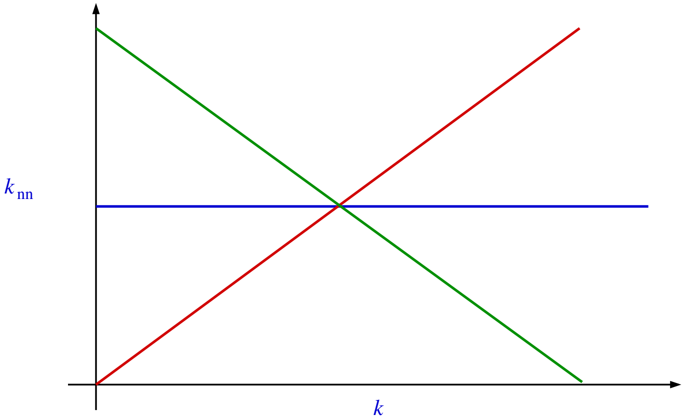
- In terms of $P(k' \mid k)$ and $e_{k'k}$:

$$k_{\rm nn}(k) = \sum_{k'} k' P(k' \mid k), \qquad P(k' \mid k) = \frac{E_{k'k}}{\sum_{k'} E_{k'k}} = \frac{e_{k'k}}{\sum_{k'} e_{k'k}}$$

$$\rightarrow \quad k_{\rm nn}(k) = \frac{\sum_{k'} k' e_{k'k}}{\sum_{k'} e_{k'k}} = \frac{\sum_{k'} k' e_{k'k}}{q_k}$$

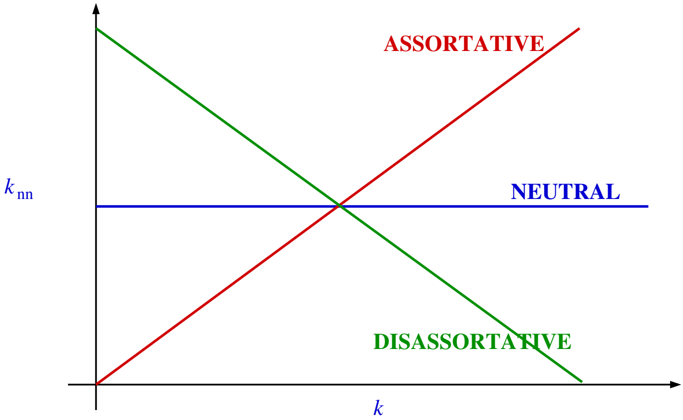
ANND

Illustration:



ANND

Illustration:



ANND

- Let's calculate the $k^{\mathrm{nn}}(k)$ in case of a neutral (un-correlated) network!

ANND

- Let's calculate the $k^{\mathrm{nn}}(k)$ in case of a neutral (un-correlated) network!

In this case $e_{k'k} = q_{k'}q_k$, thus,

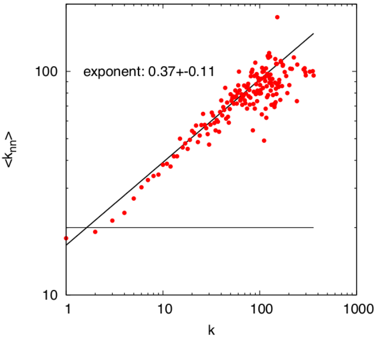
$$P(k' \mid k) = \frac{E_{k'k}}{\sum_{k'} E_{k'k}} = \frac{e_{k'k}}{\sum_{k'} e_{k'k}} = \frac{q_{k'}q_k}{q_k} = q_{k'}.$$

The ANND:

$$\begin{aligned} k_{\mathrm{nn}}(k) &= \sum_{k'} k' P(k' \mid k) = \sum_{k'} k' q_{k'} = \sum_{k'} k' \frac{k' p(k')}{\langle k \rangle} = \\ & \frac{1}{\langle k \rangle} \sum_{k'} (k')^2 p(k') = \frac{\langle k^2 \rangle}{\langle k \rangle}. \end{aligned}$$

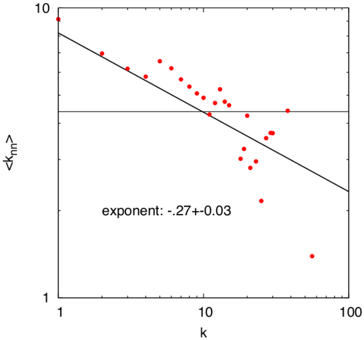
Thus, for neutral networks $k_{\mathrm{nn}}(k)$ is **independent** of k!

ANND



Astrophysics co-authorship network

Assortative



Yeast PPI

Disassortative

ANND vs $e_{k'k}$

$\mathbf{e_{k'k}}$: $k(k-1)$ parameters $\leftrightarrow$ **ANND**: k parameters

$\rightarrow$ The ANND is much simpler to evaluate and interpret.

Can we reduce the number of parameters even further?

ANND vs $e_{k'k}$

$\mathbf{e_{k'k}}$:
$k(k-1)$ parameters

$\leftrightarrow$

ANND:
k parameters

$\rightarrow$ The ANND is much simpler to evaluate and interpret.

Can we reduce the number of parameters even further?

ANND vs $e_{k'k}$

$\mathbf{e_{k'k}}$: $k(k-1)$ parameters $\leftrightarrow$ **ANND**: k parameters

$\rightarrow$ The ANND is much simpler to evaluate and interpret.

Can we reduce the number of parameters even further?

Pearson-correlation

- $e_{k'k}$ can be treated as a **joint probability distribution** for k' and k.

→ Thus, we can use the standard formulas of **co-variance** and **Pearson-correlation** to measure their relatedness.

- The co-variance:

$$\begin{aligned}
\mathrm{Cov}_e(k', k) &= \left\langle k'k \right\rangle_e - \left\langle k' \right\rangle_e \left\langle k \right\rangle_e = \\
&\sum_{k',k} k'ke_{k'k} - \left(\sum_{k,k'} k' e_{k'k}\right)\left(\sum_{k',k} ke_{k'k}\right) = \\
&\sum_{k',k} k'ke_{k'k} - \left(\sum_{k'} k' q_{k'}\right)\left(\sum_{k} kq_k\right) = \\
&\sum_{k',k} k'k\left(e_{k'k} - q_{k'}q_k\right)
\end{aligned}$$

Pearson-correlation

- $e_{k'k}$ can be treated as a **joint probability distribution** for k' and k.

→ Thus, we can use the standard formulas of **co-variance** and **Pearson-correlation** to measure their relatedness.

- The co-variance:

$$\begin{aligned}
\mathrm{Cov}_e(k',k) &= \langle k'k\rangle_e - \langle k'\rangle_e \langle k\rangle_e = \\
& \sum_{k',k} k'ke_{k'k} - \left(\sum_{k,k'} k'e_{k'k}\right)\left(\sum_{k',k} ke_{k'k}\right) = \\
& \sum_{k',k} k'ke_{k'k} - \left(\sum_{k'} k'q_{k'}\right)\left(\sum_{k} kq_{k}\right) = \\
& \sum_{k',k} k'k\left(e_{k'k} - q_{k'}q_k\right)
\end{aligned}$$

Pearson-correlation

- $e_{k'k}$ can be treated as a **joint probability distribution** for k' and k.

→ Thus, we can use the standard formulas of **co-variance** and **Pearson-correlation** to measure their relatedness.

- The co-variance:

$$\begin{aligned}\mathrm{Cov}_e(k',k) &= \left\langle k'k\right\rangle_e - \left\langle k'\right\rangle_e \left\langle k\right\rangle_e = \\ & \sum_{k',k} k'ke_{k'k} - \left(\sum_{k,k'} k'e_{k'k}\right)\left(\sum_{k',k} ke_{k'k}\right) = \\ & \sum_{k',k} k'ke_{k'k} - \left(\sum_{k'} k'q_{k'}\right)\left(\sum_{k} kq_k\right) = \\ & \sum_{k',k} k'k\left(e_{k'k} - q_{k'}q_k\right)\end{aligned}$$

Pearson-correlation

- The variance:

$$\sigma_e^2(k) = \sigma_e^2(k') = \left\langle k^2 \right\rangle_e - \langle k \rangle_e^2 = \sum_{k',k} k^2 e_{k'k} - \left(\sum_{k',k} k e_{k'k} \right)^2 = \sum_k k^2 q_k - \left(\sum_k k q_k \right)^2 .$$

- **The Pearson-correlation**:

$$r = \frac{\mathrm{Cov}_e(k',k)}{\sigma_e(k')\sigma_e(k)} = \frac{\sum_{k',k} k'k(e_{k'k} - q_{k'}q_k)}{\sigma_e^2(k)} .$$

Properties:

$$-1 \le r \le 1, \qquad \begin{array}{ll} r > 0 & \textbf{assortative} \\ r = 0 & \textbf{neutral} \\ r < 0 & \textbf{disassortative} \end{array}$$

Pearson-correlation

- The variance:

$$\sigma_e^2(k) = \sigma_e^2(k') = \left\langle k^2 \right\rangle_e - \langle k \rangle_e^2 = \sum_{k',k} k^2 e_{k'k} - \left(\sum_{k',k} k e_{k'k} \right)^2 = \sum_k k^2 q_k - \left(\sum_k k q_k \right)^2 .$$

- **The Pearson-correlation**:

$$r = \frac{\mathrm{Cov}_e(k',k)}{\sigma_e(k')\sigma_e(k)} = \frac{\sum_{k',k} k'k(e_{k'k} - q_{k'}q_k)}{\sigma_e^2(k)} .$$

Properties:

$$-1 \le r \le 1,$$

$r > 0$ **assortative**
$r = 0$ **neutral**
$r < 0$ **disassortative**

Pearson-correlation

- The variance:

$$\sigma_e^2(k) = \sigma_e^2(k') = \langle k^2 \rangle_e - \langle k \rangle_e^2 = \sum_{k',k} k^2 e_{k'k} - \left(\sum_{k',k} k e_{k'k} \right)^2 = \sum_k k^2 q_k - \left(\sum_k k q_k \right)^2 .$$

- **The Pearson-correlation**:

$$r = \frac{\mathrm{Cov}_e(k',k)}{\sigma_e(k')\sigma_e(k)} = \frac{\sum_{k',k} k'k(e_{k'k} - q_{k'}q_k)}{\sigma_e^2(k)} .$$

Properties:

$$-1 \le r \le 1, \qquad \begin{array}{ll} r > 0 & \textbf{assortative} \\ r = 0 & \textbf{neutral} \\ r < 0 & \textbf{disassortative} \end{array}$$

Pearson-correlation

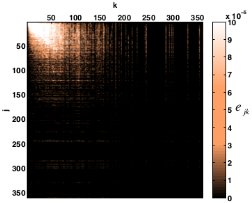
Social networks are ***assortative***

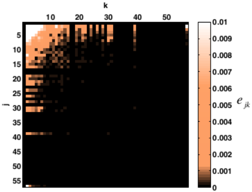
Network	n	r
Physics coauthorship (a)	52 909	0.363
Biology coauthorship (a)	1 520 251	0.127
Mathematics coauthorship (b)	253 339	0.120
Film actor collaborations (c)	449 913	0.208
Company directors (d)	7 673	0.276
Internet (e)	10 697	-0.189
World-Wide Web (f)	269 504	-0.065
Protein interactions (g)	2 115	-0.156
Neural network (h)	307	-0.163
Marine food web (i)	134	-0.247
Freshwater food web (j)	92	-0.276
Random graph (u)		0
Callaway *et al.* (v)		$\delta/(1+2\delta)$
Barabási and Albert (w)		0

Biological, technological networks are ***disassortative***

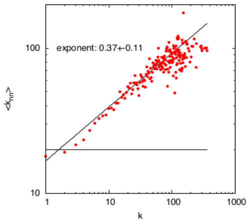
Measuring degree correlations
Summary

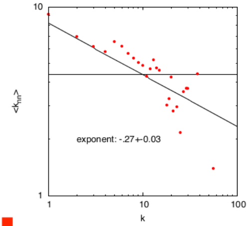
e_{jk}





$k_{annd}(k)$





r

0.31

-0.16

Network characteristics

single node	whole network		
degree, k_i strength, s_i	average degree, $\langle k \rangle$	$\leftrightarrow$	**SPARSE!**
	degree dist.,$p(k)$	$\leftrightarrow$	**SCALE-FREE!**
	ANND, $k_{\mathrm{nn}}(k)$		
distance, l_i	average distance, $\langle \ell \rangle$, diameter	$\leftrightarrow$	**SMALL WORLD!**
closeness betweenness			
clust. coeff., C	av. clustering, $\langle C \rangle$	$\leftrightarrow$	**HIGH CLUSTERING!**