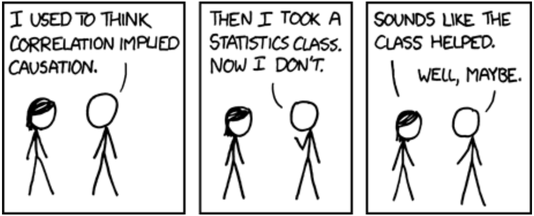
Network Science

2019. szeptember 30.

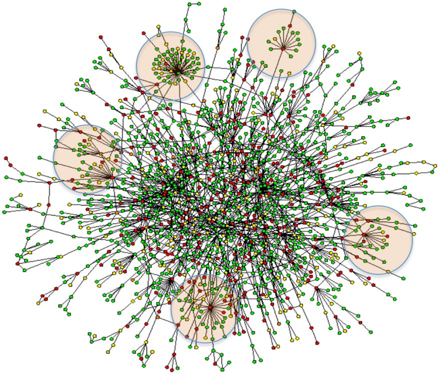


DEGREE CORRELATIONS

Disassortative „mixing” in PPI networks

- Hubs tend to link to small degree nodes in PPI networks...

→ What is the probability for having a link between nodes of degree k_i and k_j in a random graph?



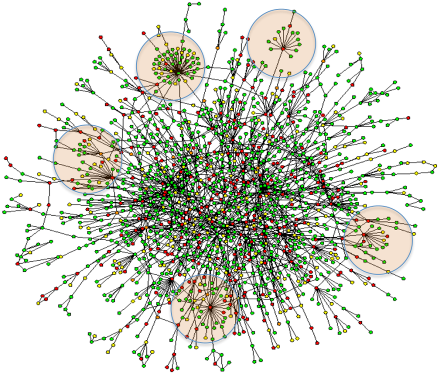
→ If $k_i = 50$, $k_j = 13$, $M = 1746$, we have $p_{50,13} = 0.15 \quad \leftrightarrow \quad p_{2,1} = 0.0004$

Yet, we see many links between degree 2 and 1 nodes, and no links between the hubs...

Disassortative „mixing” in PPI networks

- Hubs tend to link to small degree nodes in PPI networks...

→ What is the probability for having a link between nodes of degree k_i and k_j in a random graph?



→ If $k_i = 50$, $k_j = 13$, $M = 1746$, we have $p_{50,13} = 0.15 \quad \leftrightarrow \quad p_{2,1} = 0.0004$

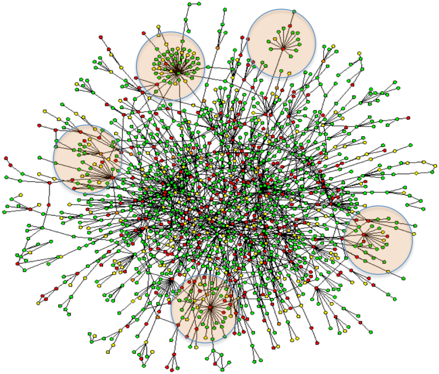
Yet, we see many links between degree 2 and 1 nodes, and no links between the hubs...

Disassortative „mixing” in PPI networks

- Hubs tend to link to small degree nodes in PPI networks...

→ What is the probability for having a link between nodes of degree k_i and k_j in a random graph?

$$P(\text{link } i - j) = \frac{k_i k_j}{2M}$$



→ If $k_i = 50$, $k_j = 13$, $M = 1746$, we have
$p_{50,13} = 0.15 \quad \leftrightarrow \quad p_{2,1} = 0.0004$

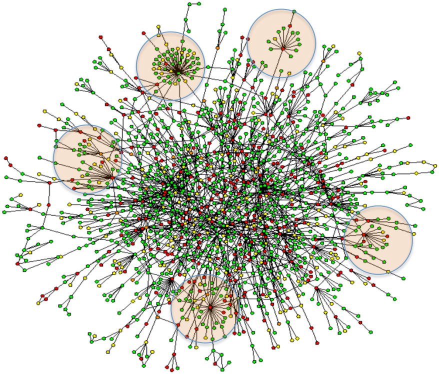
Yet, we see many links between degree 2 and 1 nodes, and no links between the hubs...

Disassortative „mixing" in PPI networks

- Hubs tend to link to small degree nodes in PPI networks...

→ What is the probability for having a link between nodes of degree k_i and k_j in a random graph?

$$P(\text{link } i-j) = \frac{k_i k_j}{2M}$$



→ If $k_i = 50$, $k_j = 13$, $M = 1746$, we have $p_{50,13} = 0.15 \quad \leftrightarrow \quad p_{2,1} = 0.0004$

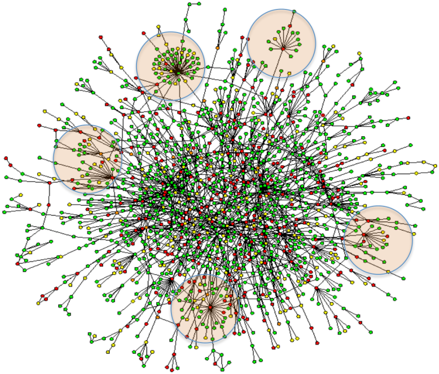
Yet, we see many links between degree 2 and 1 nodes, and no links between the hubs...

Disassortative „mixing” in PPI networks

- Hubs tend to link to small degree nodes in PPI networks...

→ What is the probability for having a link between nodes of degree k_i and k_j in a random graph?

$$P(\text{link } i-j) = \frac{k_i k_j}{2M}$$



→ If $k_i = 50$, $k_j = 13$, $M = 1746$, we have
$p_{50,13} = 0.15 \quad \leftrightarrow \quad p_{2,1} = 0.0004$

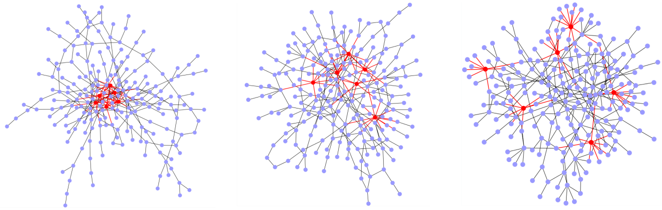
Yet, we see many links between degree 2 and 1 nodes, and no links between the hubs...

Assortative and disassortative networks

Assortativity and disassortativity

- **Assortative network**: small degree nodes tend to connect to other small degree nodes, hubs tend to link to each other.
- Neutral network: nodes connect to each other at random.
- **Disasortative network**: hubs avoid linking to each other, instead they connect to small degree nodes.

Illustration:



assortative neutral dissasortative

How to describe assortativity?

How to describe assortativity?

Full statistical description:

- Def.: let $P(k' \mid k)$ denote the **conditional probability** for finding a **node with degree** k' at one end of a link, **given** the node at the other end has **degree** k.
- In principle, $P(k' \mid k)$ encodes all info about whether the network is assortative or disassortative.
- How to measure this in practice?
  By definition:

$$P(k' \mid k) = \frac{P(\text{link between } k' \text{ and } k)}{P(\text{link on } k)}.$$

→ Def.: let $E_{k',k}$ count the number of links between nodes of degree k' and k, and

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

How to describe assortativity?

Full statistical description:

- Def.: let $P(k' \mid k)$ denote the **conditional probability** for finding a **node with degree** k' at one end of a link, **given** the node at the other end has **degree** k.
- In principle, $P(k' \mid k)$ encodes all info about whether the network is assortative or disassortative.
- How to measure this in practice?
  By definition:

$$P(k' \mid k) = \frac{P(\text{link between } k' \text{ and } k)}{P(\text{link on } k)}.$$

→ Def.: let $E_{k',k}$ count the number of links between nodes of degree k' and k, and

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

How to describe assortativity?

Full statistical description:

- Def.: let $P(k' \mid k)$ denote the **conditional probability** for finding a **node with degree** k' at one end of a link, **given** the node at the other end has **degree** k.
- In principle, $P(k' \mid k)$ encodes all info about whether the network is assortative or disassortative.
- How to measure this in practice?
  By definition:

$$P(k' \mid k) = \frac{P(\text{link between } k' \text{ and } k)}{P(\text{link on } k)}.$$

→ Def.: let $E_{k',k}$ count the number of links between nodes of degree k' and k, and

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

How to describe assortativity?

Full statistical description:

- Def.: let $P(k' \mid k)$ denote the **conditional probability** for finding a **node with degree** k' at one end of a link, **given** the node at the other end has **degree** k.
- In principle, $P(k' \mid k)$ encodes all info about whether the network is assortative or disassortative.
- How to measure this in practice?
  By definition:

$$P(k' \mid k) = \frac{P(\text{link between } k' \text{ and } k)}{P(\text{link on } k)}.$$

→ Def.: let $E_{k',k}$ count the number of links between nodes of degree k' and k, and

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

How to describe assortativity?

Full statistical description:

- Def.: let $P(k' \mid k)$ denote the **conditional probability** for finding a **node with degree** k' at one end of a link, **given** the node at the other end has **degree** k.
- In principle, $P(k' \mid k)$ encodes all info about whether the network is assortative or disassortative.
- How to measure this in practice?
  By definition:

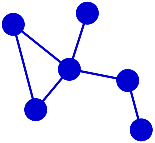
$$P(k' \mid k) = \frac{P(\text{link between } k' \text{ and } k)}{P(\text{link on } k)}.$$

→ Def.: let $E_{k',k}$ count the number of links between nodes of degree k' and k, and

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

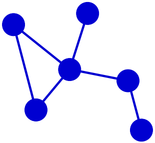
Full statistical description

- $E_{k',k}$: number of links between nodes of degree k' and k, links between nodes with the same degree count twice!



Full statistical description

- $E_{k',k}$: number of links between nodes of degree k' and k, links between nodes with the same degree count twice!

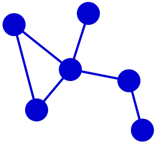


$\rightarrow E_{k',k}:$

$k=$	1	2	3	4
1	0	1	0	1
2	1	2	0	3
3	0	0	0	0
4	1	3	0	0

Full statistical description

- $E_{k',k}$: number of links between nodes of degree k' and k, links between nodes with the same degree count twice!



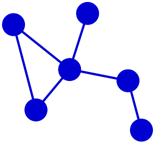
$\rightarrow E_{k',k}$:

$k=$	1	2	3	4
1	0	1	0	1
2	1	2	0	3
3	0	0	0	0
4	1	3	0	0

$\rightarrow$ If we are going to measure $E_{k',k}$, we might as well „forget” $P(k' \mid k)$, and examine what does assortativity mean in turns of $E_{k',k}$.

Full statistical description

- $E_{k',k}$: number of links between nodes of degree k' and k, links between nodes with the same degree count twice!



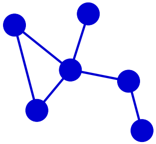
$\rightarrow E_{k',k}$:

$k =$	1	2	3	4
1	0	1	0	1
2	1	2	0	3
3	0	0	0	0
4	1	3	0	0

- Def.: $e_{k',k} = \frac{E_{k',k}}{2M}$ gives 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.

Full statistical description

- $E_{k',k}$: number of links between nodes of degree k' and k, links between nodes with the same degree count twice!



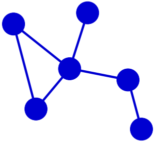
$\rightarrow e_{k',k}$:

$k =$	1	2	3	4
1	0	$\frac{1}{12}$	0	$\frac{1}{12}$
2	$\frac{1}{12}$	$\frac{1}{6}$	0	$\frac{1}{4}$
3	0	0	0	0
4	$\frac{1}{12}$	$\frac{1}{4}$	0	0

- Def.: $e_{k',k} = \frac{E_{k',k}}{2M}$ gives 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.

Full statistical description

- $E_{k',k}$: number of links between nodes of degree k' and k, links between nodes with the same degree count twice!



$\rightarrow e_{k',k}$:

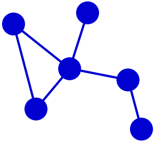
$k =$	1	2	3	4
1	0	$\frac{1}{12}$	0	$\frac{1}{12}$
2	$\frac{1}{12}$	$\frac{1}{6}$	0	$\frac{1}{4}$
3	0	0	0	0
4	$\frac{1}{12}$	$\frac{1}{4}$	0	0

- Def.: $e_{k',k} = \frac{E_{k',k}}{2M}$ gives 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.

$$\sum_{k'k} e_{k'k} = 1$$

Full statistical description

- $E_{k',k}$: number of links between nodes of degree k' and k, links between nodes with the same degree count twice!



$\rightarrow e_{k',k}$:

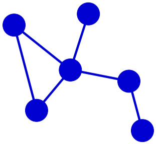
$k =$	1	2	3	4
1	0	$\frac{1}{12}$	0	$\frac{1}{12}$
2	$\frac{1}{12}$	$\frac{1}{6}$	0	$\frac{1}{4}$
3	0	0	0	0
4	$\frac{1}{12}$	$\frac{1}{4}$	0	0

- Def.: $e_{k',k} = \frac{E_{k',k}}{2M}$ gives 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.

$$\sum_{k'k} e_{k'k} = 1 \qquad \sum_{k'} e_{k'k} \overset{\text{Def.}}{=} q_k.$$

Full statistical description

- $E_{k',k}$: number of links between nodes of degree k' and k, links between nodes with the same degree count twice!



$\rightarrow e_{k',k}$:

$k =$	1	2	3	4
1	0	$\frac{1}{12}$	0	$\frac{1}{12}$
2	$\frac{1}{12}$	$\frac{1}{6}$	0	$\frac{1}{4}$
3	0	0	0	0
4	$\frac{1}{12}$	$\frac{1}{4}$	0	0

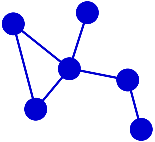
- Def.: $e_{k',k} = \frac{E_{k',k}}{2M}$ gives the probability for finding a node with degree k' at one end and a node with degree k at the other end of a randomly selected link.

$$\sum_{k'k} e_{k'k} = 1 \qquad \sum_{k'} e_{k'k} \overset{\text{Def.}}{=} q_k .$$

What is q_k?

Full statistical description

- $E_{k',k}$: number of links between nodes of degree k' and k, links between nodes with the same degree count twice!



$\rightarrow e_{k',k}$:

$k =$	1	2	3	4
1	0	$\frac{1}{12}$	0	$\frac{1}{12}$
2	$\frac{1}{12}$	$\frac{1}{6}$	0	$\frac{1}{4}$
3	0	0	0	0
4	$\frac{1}{12}$	$\frac{1}{4}$	0	0
q_k	$\frac{1}{6}$	$\frac{1}{2}$	0	$\frac{1}{3}$

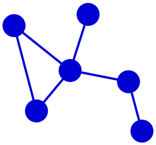
- Def.: $e_{k',k} = \frac{E_{k',k}}{2M}$ gives the probability for finding a node with degree k' at one end and a node with degree k at the other end of a randomly selected link.

$$\sum_{k'k} e_{k'k} = 1 \qquad\qquad \sum_{k'} e_{k'k} \overset{\text{Def.}}{=} q_k.$$

What is q_k?

Full statistical description

- $E_{k',k}$: number of links between nodes of degree k' and k, links between nodes with the same degree count twice!



$\rightarrow e_{k',k}$:

$k =$	1	2	3	4
1	0	$\frac{1}{12}$	0	$\frac{1}{12}$
2	$\frac{1}{12}$	$\frac{1}{6}$	0	$\frac{1}{4}$
3	0	0	0	0
4	$\frac{1}{12}$	$\frac{1}{4}$	0	0
q_k	$\frac{1}{6}$	$\frac{1}{2}$	0	$\frac{1}{3}$

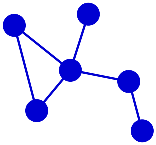
- Def.: $e_{k',k} = \frac{E_{k',k}}{2M}$ gives the probability for finding a node with degree k' at one end and a node with degree k at the other end of a randomly selected link.

$$\sum_{k'k} e_{k'k} = 1 \qquad\qquad \sum_{k'} e_{k'k} \overset{\text{Def.}}{=} q_k .$$

What is q_k? $\rightarrow$ the probability for finding a node with degree k at one end of a randomly selected link.

Full statistical description

- $E_{k',k}$: number of links between nodes of degree k' and k, links between nodes with the same degree count twice!



$\rightarrow e_{k',k}$:

$k =$	1	2	3	4
1	0	$\frac{1}{12}$	0	$\frac{1}{12}$
2	$\frac{1}{12}$	$\frac{1}{6}$	0	$\frac{1}{4}$
3	0	0	0	0
4	$\frac{1}{12}$	$\frac{1}{4}$	0	0
q_k	$\frac{1}{6}$	$\frac{1}{2}$	0	$\frac{1}{3}$

- Def.: $e_{k',k} = \frac{E_{k',k}}{2M}$ gives 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.

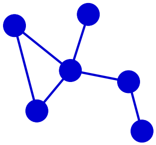
$$\sum_{k'k} e_{k'k} = 1 \qquad\qquad \sum_{k'} e_{k'k} \stackrel{\text{Def.}}{=} q_k.$$

What is q_k? $\rightarrow$ the probability for finding a node with degree k at one end of a randomly selected link.

$\rightarrow$ In a neutral network with no degree correlations: $e_{k'k} = q_{k'} \cdot q_k$.

Full statistical description

- $E_{k',k}$: number of links between nodes of degree k' and k, links between nodes with the same degree count twice!



$\rightarrow e_{k',k}$:

$k =$	1	2	3	4
1	0	$\frac{1}{12}$	0	$\frac{1}{12}$
2	$\frac{1}{12}$	$\frac{1}{6}$	0	$\frac{1}{4}$
3	0	0	0	0
4	$\frac{1}{12}$	$\frac{1}{4}$	0	0
q_k	$\frac{1}{6}$	$\frac{1}{2}$	0	$\frac{1}{3}$

- Def.: $e_{k',k} = \frac{E_{k',k}}{2M}$ gives 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.

$$\sum_{k'k} e_{k'k} = 1 \qquad \sum_{k'} e_{k'k} \overset{\text{Def.}}{=} q_k.$$

What is q_k? $\rightarrow$ the probability for finding a node with degree k at one end of a randomly selected link.

$\rightarrow$ In a neutral network with no degree correlations: $e_{k'k} = q_{k'} \cdot q_k$.

Thus, the **deviations from this value are the signatures 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:
  $$q_k = \frac{kN_k}{\sum_{k'} k'N_{k'}} = \frac{kp(k)N}{\sum_{k'} k'p(k')N} = \frac{kp(k)}{\langle k \rangle}$$

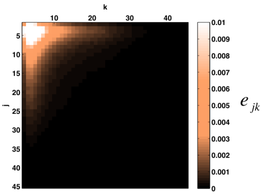
Full statistical description

Full statistical description of assortativity

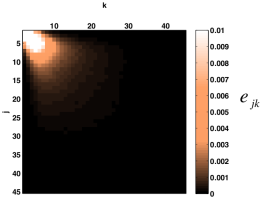
- $E_{k',k}$: number of links between nodes of degree k' and k.
- $e_{k',k} = \frac{E_{k',k}}{2M}$: the probability for finding a node with degree k' at one end and a node with degree k at the other end of a randomly selected link,
- $q_k = \frac{kp(k)}{\langle k \rangle}$: prob. of degree k at one end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

Statistical description of assortativity

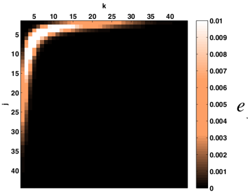
Averaging on 100 samples of SF networks with
$N = 10^4$, $\gamma = 2.5$, $\langle k \rangle = 3$:



Neutral



Assortative

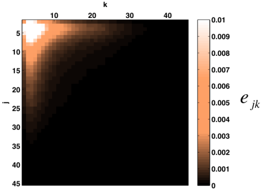


Disassortative

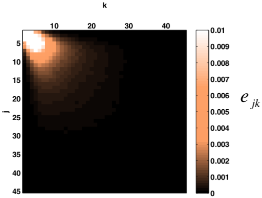
How would the distribution look like for a perfectly assortative/disassortative network?

Statistical description of assortativity

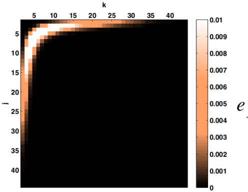
Averaging on 100 samples of SF networks with $N = 10^4$, $\gamma = 2.5$, $\langle k \rangle = 3$:



Neutral



Assortative

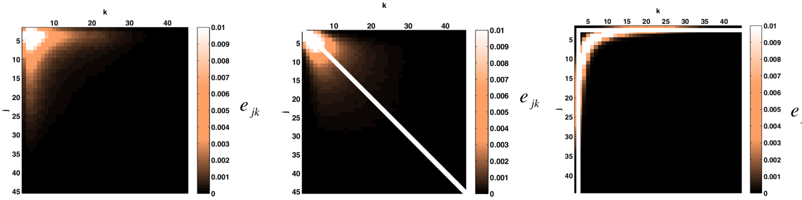


Disassortative

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

Statistical description of assortativity

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

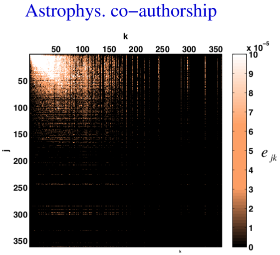


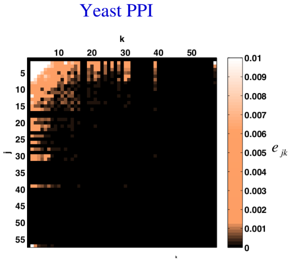
Neutral Assortative Disassortative

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

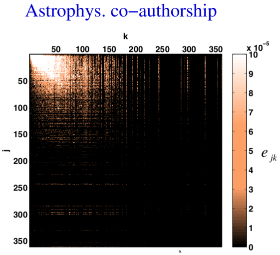
Statistical description of assortativity

Results for real networks:

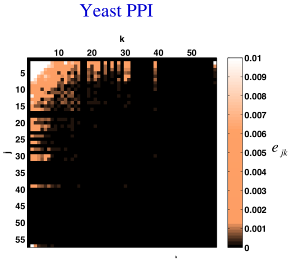




Results for real networks:



assortative



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 of individual nodes

- The ANND of node i:

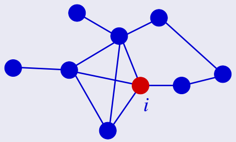
$$k_i^{\mathrm{ANND}} \equiv \langle k_j \rangle_{j \text{ linked to } i} = \frac{1}{k_i} \sum_{j \text{ linked to } i} k_j .$$

ANND

ANND of individual nodes

- The ANND of node i:

$$k_i^{\mathrm{ANND}} \equiv \langle k_j \rangle_{j \text{ linked to } i} = \frac{1}{k_i} \sum_{j \text{ linked to } i} k_j .$$



$$k_i^{\mathrm{ANND}} = \frac{2+5+4+3}{4} = 3.5$$

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:

$$k^{\mathrm{ANND}}(k) \equiv \left\langle k_i^{\mathrm{ANND}} \right\rangle_{k_i=k}.$$

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

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

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

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:

$$k^{\mathrm{ANND}}(k) \equiv \left\langle k_i^{\mathrm{ANND}} \right\rangle_{k_i=k}.$$

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

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

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

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:

$$k^{\mathrm{ANND}}(k) \equiv \left\langle k_i^{\mathrm{ANND}} \right\rangle_{k_i=k}.$$

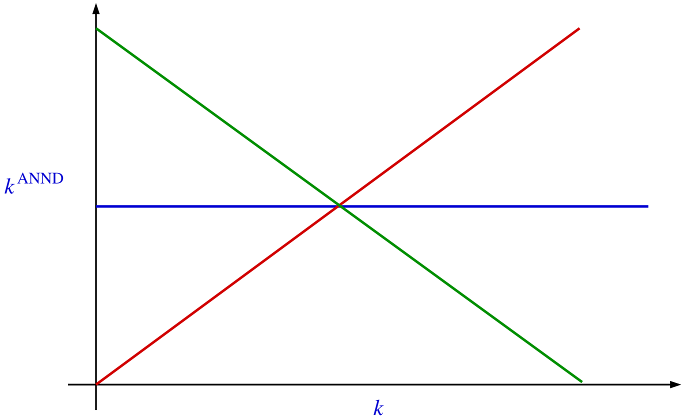
- In terms of $P(k' \mid k)$ and $e_{k'k}$:

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

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

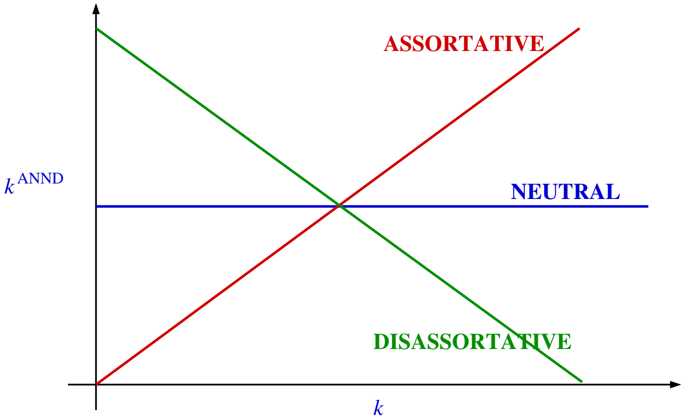
ANND

Illustration:



ANND

Illustration:



ANND

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

ANND

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

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

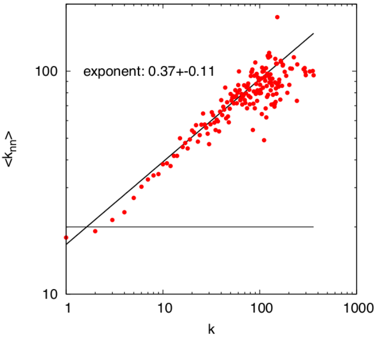
$$P(k' \mid k) = \frac{E_{k'k}}{\sum_{k'} E_{k'k}} = \frac{e_{k'k}}{\sum_{k'} e_{k'k}} = \frac{q_{k'}q_k}{q_k} = q_{k'}.$$

The ANND:

$$k^{\mathrm{ANND}}(k) = \sum_{k'} k' P(k' \mid k) = \sum_{k'} k' q_{k'} = \sum_{k'} k' \frac{k' p(k')}{\langle k \rangle} = \frac{1}{\langle k \rangle} \sum_{k'} (k')^2 p(k') = \frac{\langle k^2 \rangle}{\langle k \rangle}.$$

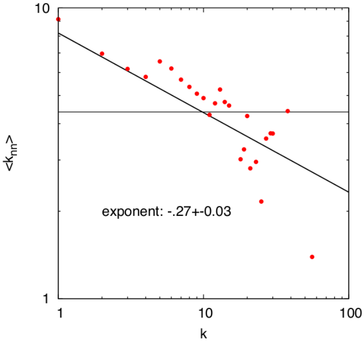
Thus, for neutral networks $k^{\mathrm{ANND}}(k)$ is **independent** of k!

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.

$\rightarrow$ 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:

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

Pearson-correlation

- The variance:

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

- **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

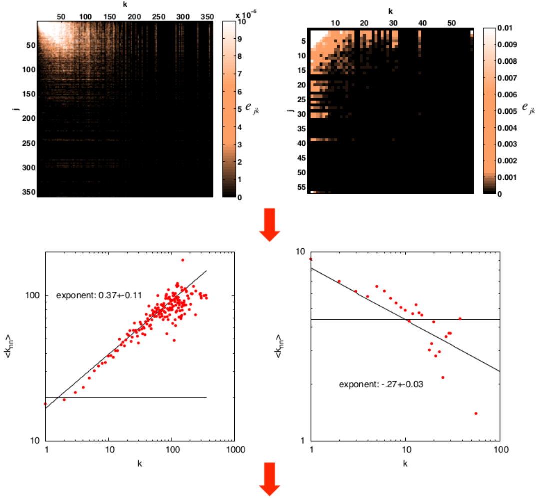
Network	n	r
Physics coauthorship (a)	52 909	0.363
Biology coauthorship (a)	1 520 251	0.127
Mathematics coauthorship (b)	253 339	0.120
Film actor collaborations (c)	449 913	0.208
Company directors (d)	7 673	0.276
Internet (e)	10 697	−0.189
World-Wide Web (f)	269 504	−0.065
Protein interactions (g)	2 115	−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

Social networks are ***assortative***

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, $\langle k \rangle$ degree dist.,$p(k)$ ANND, $k^{\mathrm{ANND}}(k)$	$\leftrightarrow$ $\leftrightarrow$	**SPARSE!** **SCALE-FREE!**
distance, l_i	average distance, $\langle l \rangle$, diameter	$\leftrightarrow$	**SMALL WORLD!**
closeness betweenness			
clust. coeff., C	av. clustering, $\langle C \rangle$	$\leftrightarrow$	**HIGH CLUSTERING!**