# *The Blind Watchmaker Network:*
## *Scale-freeness and evolution*

Sebastian Bernhardsson

Umeå university, Sweden

*sebbeb@tp.umu.se*

*www.tp.umu.se/∼sebbeb*

# Outline

- Statistical mechanics
- Balls in boxes model
- Variational calculus
- Variational calculus using a random process
- Constrained Balls in Boxes model
- Network constraints
- Comparison with real networks
- Other network properties (C-r space)
- Conclusions

# Statistical mechanics

## Maximum entropy principle

# Statistical mechanics

## Maximum entropy principle

*In the limit of large numbers, the outcome of a random process will be the distribution that has the largest entropy.*

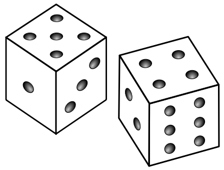
# Statistical mechanics

## Maximum entropy principle

*In the limit of large numbers, the outcome of a random process will be the distribution that has the largest entropy.*

**Example:** Flip a coin with sides **A** and **B**. How many **A**:s and **B**:s will you most likely have after four flips?

# Statistical mechanics



## Maximum entropy principle

*In the limit of large numbers, the outcome of a random process will be the distribution that has the largest entropy.*

**Example:** Flip a coin with sides **A** and **B**. How many **A**:s and **B**:s will you most likely have after four flips?

Distr. 1: 1 state { **AAAA**

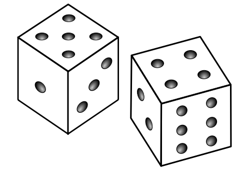
Distr. 2: 4 states { **AAAB**, **AABA**, **ABAA**, **BAAA**

Distr. 3: 6 states { **AABB**, **ABAB**, **ABBA**, **BABA**, **BAAB**, **BBAA**

Distr. 4: 4 states { **BBBA**, **BBAB**, **BABB**, **ABBB**

Distr. 5: 1 state { **BBBB**

# Statistical mechanics



## Maximum entropy principle

*In the limit of large numbers, the outcome of a random process will be the distribution that has the largest entropy.*

**Example:** Flip a coin with sides **A** and **B**. How many **A**:s and **B**:s will you most likely have after four flips?

**Answer:** The distribution (number of **A**:s and **B**:s) that has the most number of states (highest entropy) will win (in the long run)!

In this case distribution 3 is most likely to show up (prob = 6/(1+4+6+4+1) = 3/8).

Distr. 1: 1 state { **AAAA**

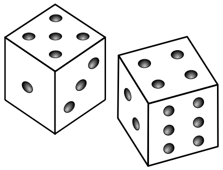
Distr. 2: 4 states { **AAAB**, **AABA**, **ABAA**, **BAAA**

Distr. 3: 6 states { **AABB**, **ABAB**, **ABBA**, **BABA**, **BAAB**, **BBAA**

Distr. 4: 4 states { **BBBA**, **BBAB**, **BABB**, **ABBB**

Distr. 5: 1 state { **BBBB**

# Network → Balls in boxes model

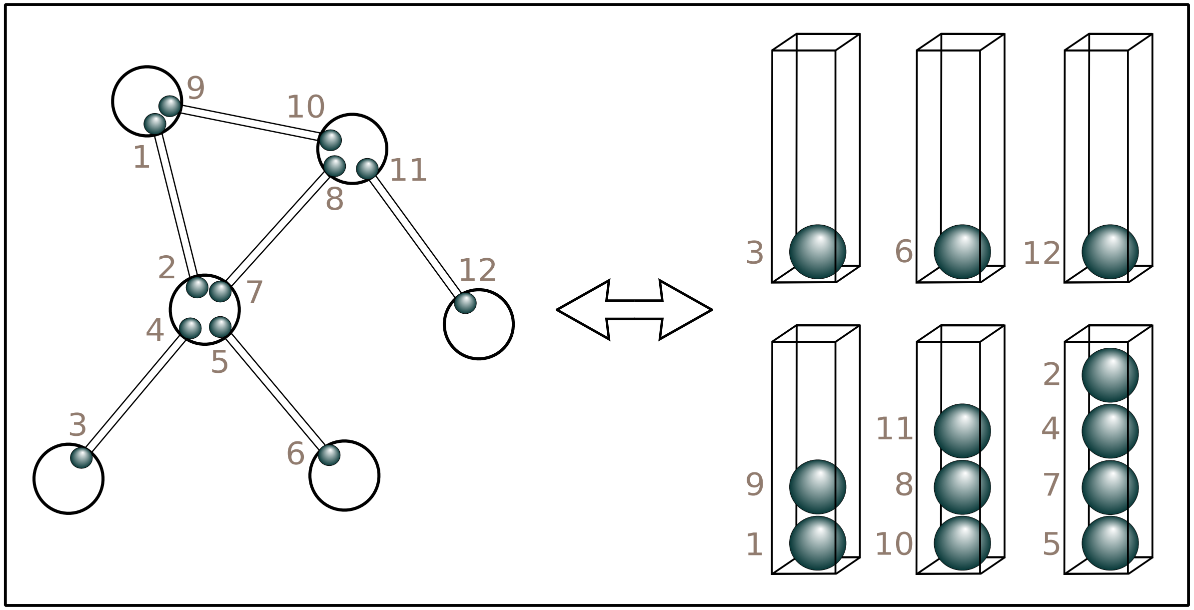


Map a network onto a set of balls and boxes.

Boxes $\Longleftrightarrow$ Nodes

Balls $\Longleftrightarrow$ Link ends

A link is defined by two link ends, e.g. (7,8)



# Variational Calculus

In how many ways can you distribute $M$ balls into $N$ boxes?

# Variational Calculus

In how many ways can you distribute $M$ balls into $N$ boxes?

$$\Omega = \frac{M!}{\prod_k N(k)! \cdot (k!)^{N(k)}} \text{ (Indistinguishable balls in a box)}$$

# Variational Calculus

In how many ways can you distribute $M$ balls into $N$ boxes?

$$\Omega = \frac{M!}{\prod_k N(k)! \cdot (k!)^{N(k)}} \text{ (Indistinguishable balls in a box)}$$

$$\Rightarrow \ln \Omega \approx M \ln M - M - \sum_k N(k)[\ln N(k) - 1] - \sum_k N(k) \ln k!$$

# Variational Calculus

In how many ways can you distribute $M$ balls into $N$ boxes?

$$\Omega = \frac{M!}{\prod_k N(k)! \cdot (k!)^{N(k)}} \text{ (Indistinguishable balls in a box)}$$

$$\Rightarrow \ln \Omega \approx M \ln M - M - \sum_k N(k)[\ln N(k) - 1] - \sum_k N(k) \ln k!$$

To find the maximum of $\ln \Omega$, we put $\frac{d}{dN(k)} \ln \Omega = 0$ with the constraints $\sum_k N(k) = N$ and $\sum_k N(k)k = M$.

# Variational Calculus

In how many ways can you distribute $M$ balls into $N$ boxes?

$$\Omega = \frac{M!}{\prod_k N(k)! \cdot (k!)^{N(k)}} \text{ (Indistinguishable balls in a box)}$$

$$\Rightarrow \ln\Omega \approx M \ln M - M - \sum_k N(k)[\ln N(k) - 1] - \sum_k N(k) \ln k!$$

To find the maximum of $\ln\Omega$, we put $\frac{d}{dN(k)} \ln\Omega = 0$ with the constraints $\sum_k N(k) = N$ and $\sum_k N(k)k = M$.

$$\Rightarrow 0 = -\ln N(k) + 1 - 1 - a - bk - \ln k!$$

(where $a$ and $b$ are Lagrange multipliers corresponding to the constraints)

# Variational Calculus

In how many ways can you distribute $M$ balls into $N$ boxes?

$$\Omega = \frac{M!}{\prod_k N(k)! \cdot (k!)^{N(k)}} \text{ (Indistinguishable balls in a box)}$$

$$\Rightarrow \ln \Omega \approx M \ln M - M - \sum_k N(k)[\ln N(k) - 1] - \sum_k N(k) \ln k!$$

To find the maximum of $\ln \Omega$, we put $\frac{d}{dN(k)} \ln \Omega = 0$ with the constraints $\sum_k N(k) = N$ and $\sum_k N(k) k = M$.

$$\Rightarrow 0 = -\ln N(k) + 1 - 1 - a - bk - \ln k!$$

(where $a$ and $b$ are Lagrange multipliers corresponding to the constraints)

$$\Rightarrow N(k) = Ae^{-bk}/k! = A\langle k \rangle^k / k! \Rightarrow \text{Poisson distribution (Erdős-Renyi)}$$

# VC using a random process

# VC using a random process

Mapping changes in the states to changes in the degree distribution.

# VC using a random process

Mapping changes in the states to changes in the degree distribution.

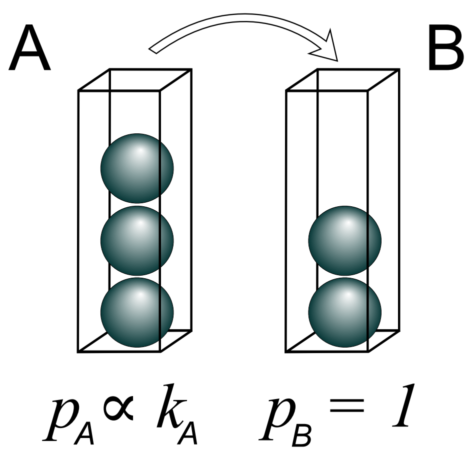
If you, by moving one ball from box $A$ to box $B$, can reach $P$ different states, then this move should get a weight $\propto P$.

# VC using a random process

Mapping changes in the states to changes in the degree distribution.

If you, by moving one ball from box $A$ to box $B$, can reach $P$ different states, then this move should get a weight $\propto P$.

There are $k_A$ balls to choose from in box A but only one place to put it in box B $\Rightarrow P = p_A \cdot p_B = k_A$.

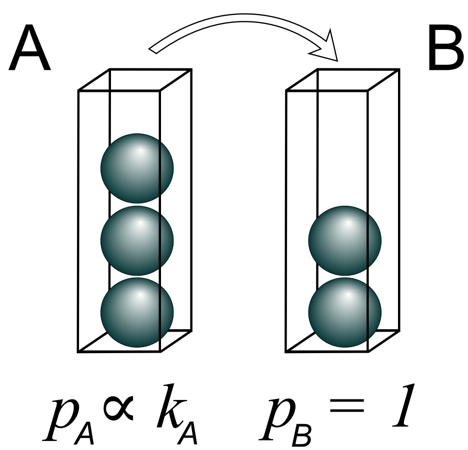


# VC using a random process

Mapping changes in the states to changes in the degree distribution.

If you, by moving one ball from box $A$ to box $B$, can reach $P$ different states, then this move should get a weight $\propto P$.

There are $k_A$ balls to choose from in box A but only one place to put it in box B
$\Rightarrow P = p_A \cdot p_B = k_A$.



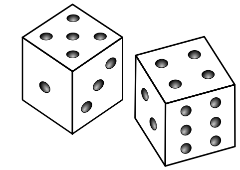

$\Longrightarrow$

**The random process**

1) Pick one box with probability $p \propto k$ and one box with $p = 1/N$.

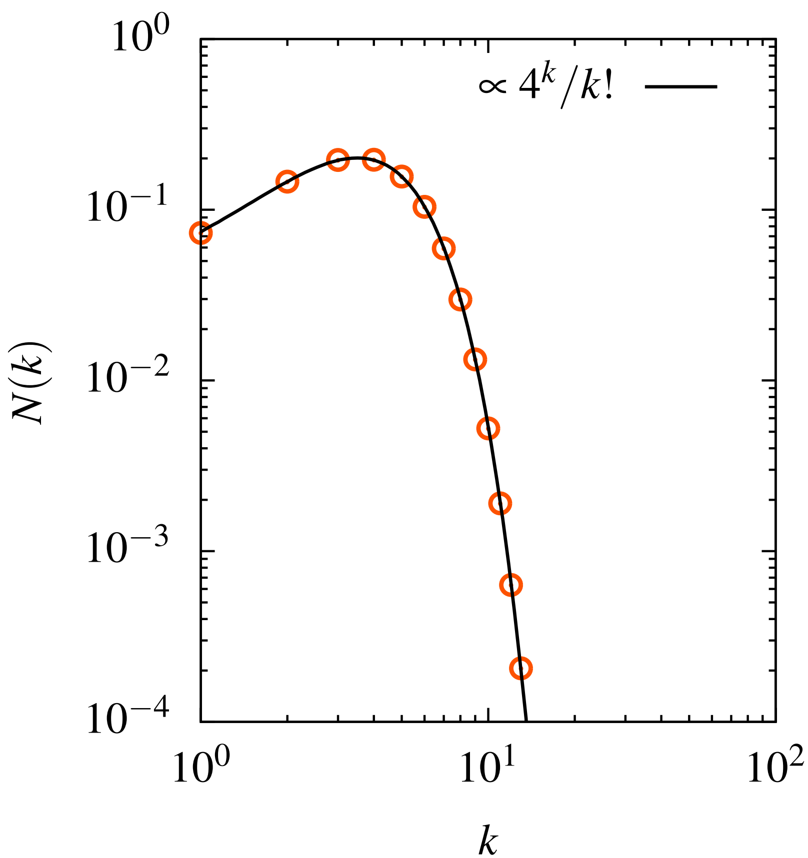
2) Move one ball from the first box to the second box.

# VC using a random process



Mapping changes in the states to changes in the degree distribution.

If you, by moving one ball from box $A$ to box $B$, can reach $P$ different states, then this move should get a weight $\propto P$.




$\Leftarrow$

**The random process**

1) Pick one box with probability $p \propto k$ and one box with $p = 1/N$.

2) Move one ball from the first box to the second box.

# Constrained Balls in Boxes

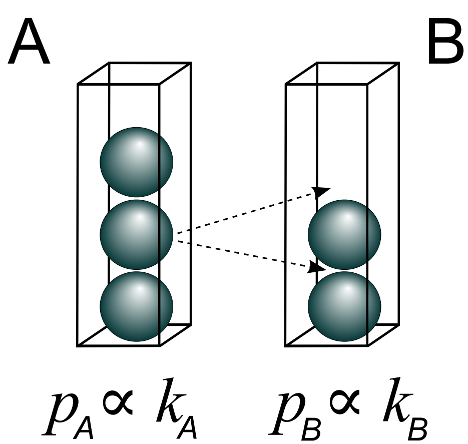
# Constrained Balls in Boxes

$k$-degeneracy (cyclic degeneracy) $\rightarrow$ One ball in each box.

# Constrained Balls in Boxes

$k$-degeneracy (cyclic degeneracy) $\rightarrow$ One ball in each box.

$\Rightarrow$ we have $k_A$ choices from box $A$ and $k_B$ choices from box $B$.

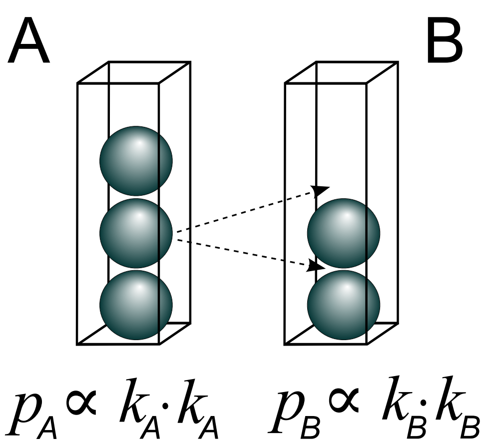


# Constrained Balls in Boxes

$k$-degeneracy (cyclic degeneracy) $\rightarrow$ One ball in each box.

$\Rightarrow$ we have $k_A$ choices from box $A$ and $k_B$ choices from box $B$.

If we also take into consideration that you can only choose a box by choosing a ball, we get an extra $k$ for each box.

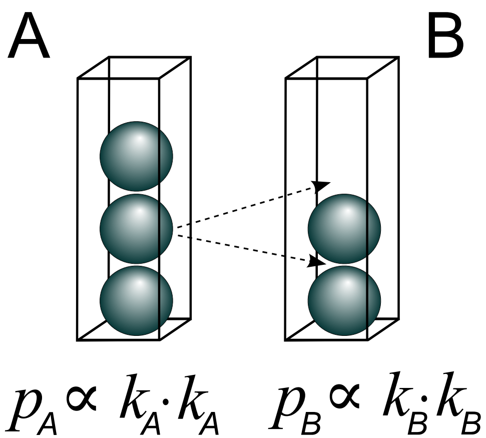


# Constrained Balls in Boxes

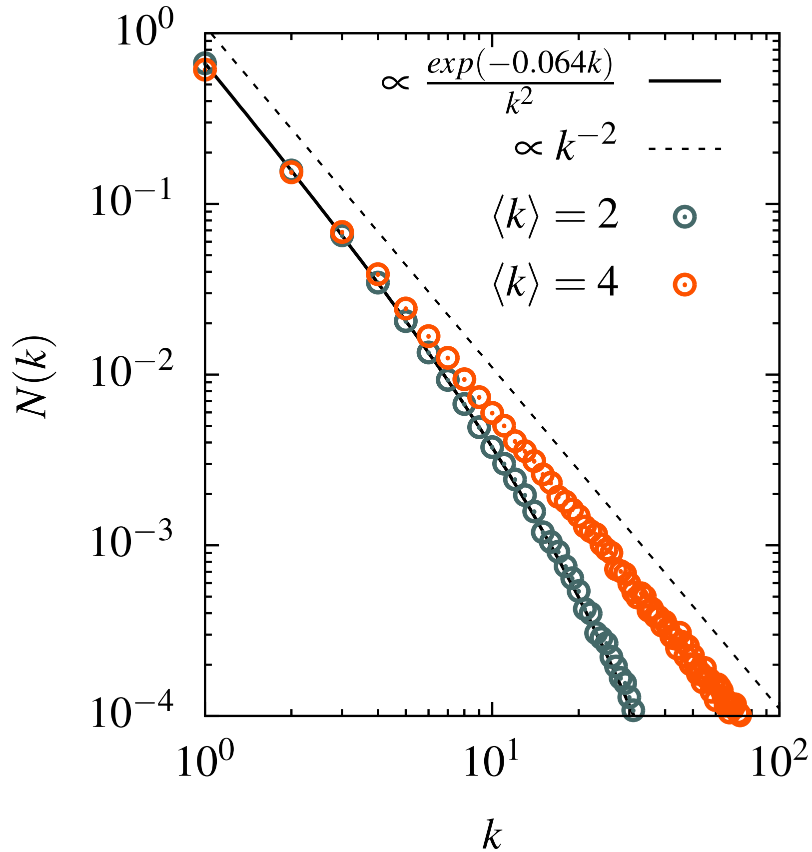
$k$-degeneracy (cyclic degeneracy) $\rightarrow$ One ball in each box.

$\Rightarrow$ we have $k_A$ choices from box $A$ and $k_B$ choices from box $B$.

If we also take into consideration that you can only choose a box by choosing a ball, we get an extra $k$ for each box.




$\Rightarrow$



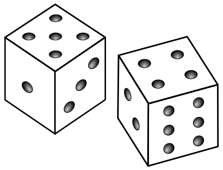

$$\Rightarrow N(k) = Ae^{-bk}/k^2$$

You are not allowed to empty a box!

# The Blind Watchmaker Network

## Network constraints

# The Blind Watchmaker Network



## Network constraints

- Only one link between two nodes

# The Blind Watchmaker Network

## Network constraints

- Only one link between two nodes
- No self loops

# The Blind Watchmaker Network

## Network constraints

- Only one link between two nodes
- No self loops

Implement the constraints in the algorithm in the following way:

1) Pick two nodes with prob, $p \propto k^2$.
2) Move a random link from one to the other. If the attempt is not allowed by network constraint try another random link from the same node. Repeat until one is moved. If none can be moved start from 1).

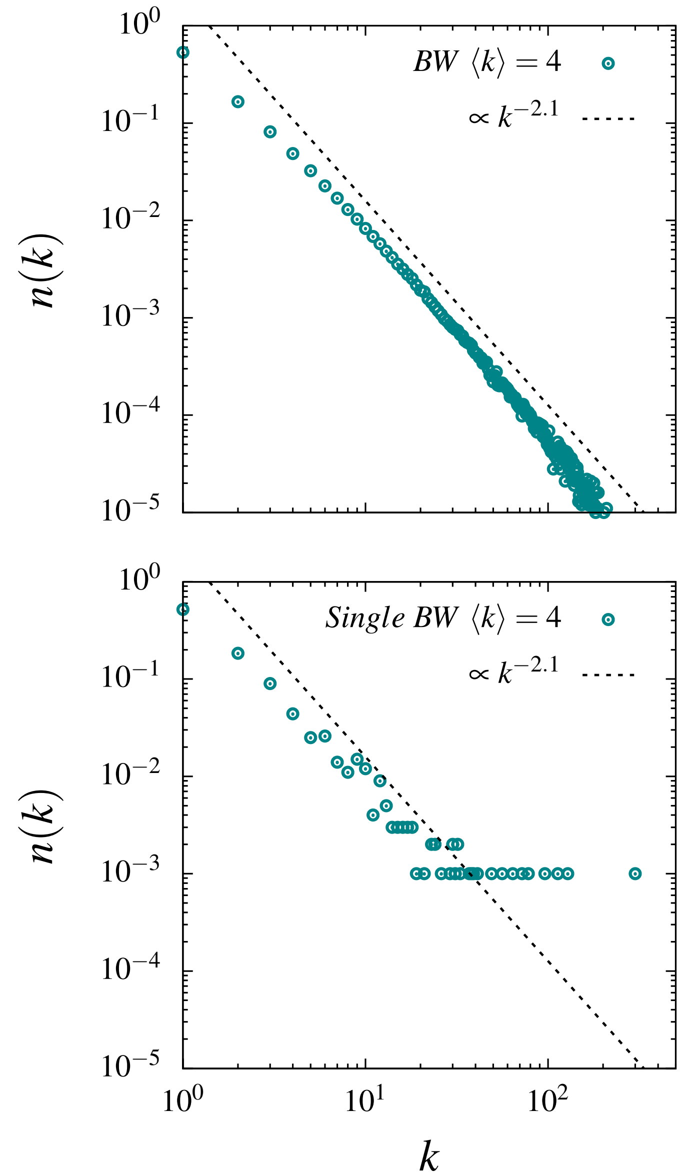
# The Blind Watchmaker Network

## Network constraints

- Only one link between two nodes
- No self loops

Implement the constraints in the algorithm in the following way:

1) Pick two nodes with prob, $p \propto k^2$.
2) Move a random link from one to the other. If the attempt is not allowed by network constraint try another random link from the same node. Repeat until one is moved. If none can be moved start from 1).



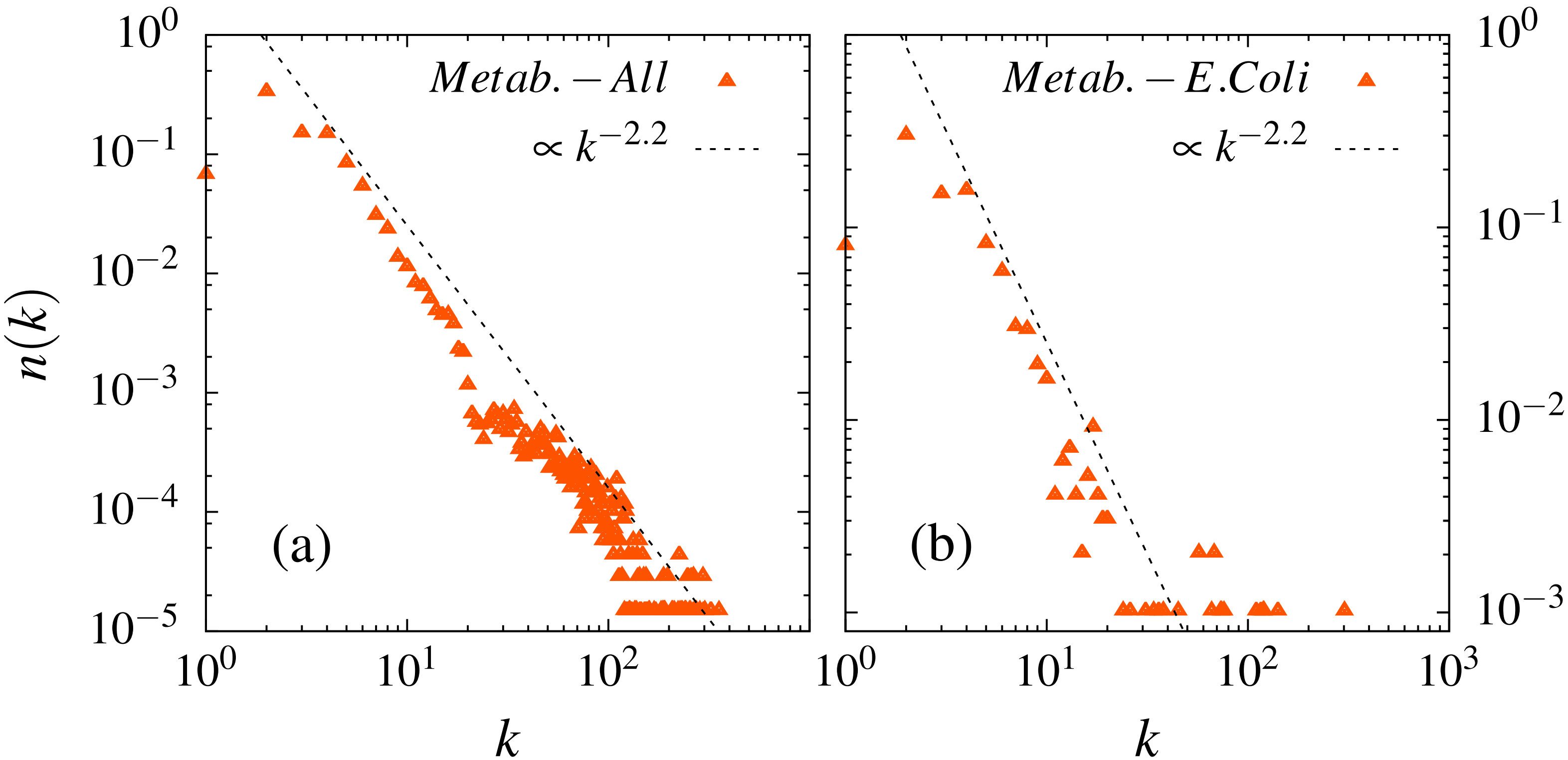
# Comparison with real data

# Comparison with real data

Metabolic networks:

(a) Average over 107 organisms

(b) E. Coli



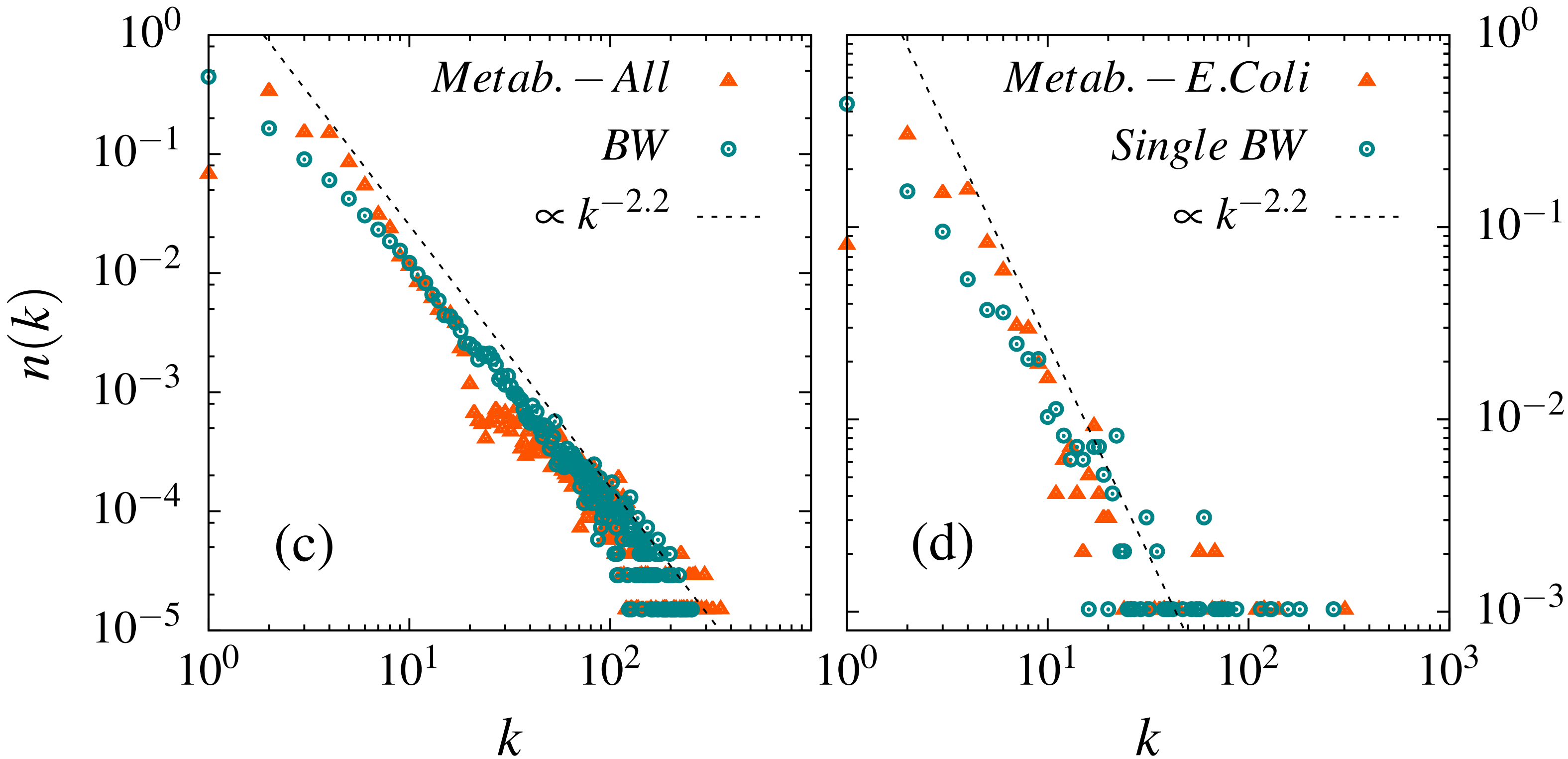

Ma H and Zeng A-P, Bioinformatics 19: 270-277 (2003).

# Comparison with real data

Comparison:

(c) BW vs Metabolic: same N and M

(d) BW vs E.Coli: same N and M

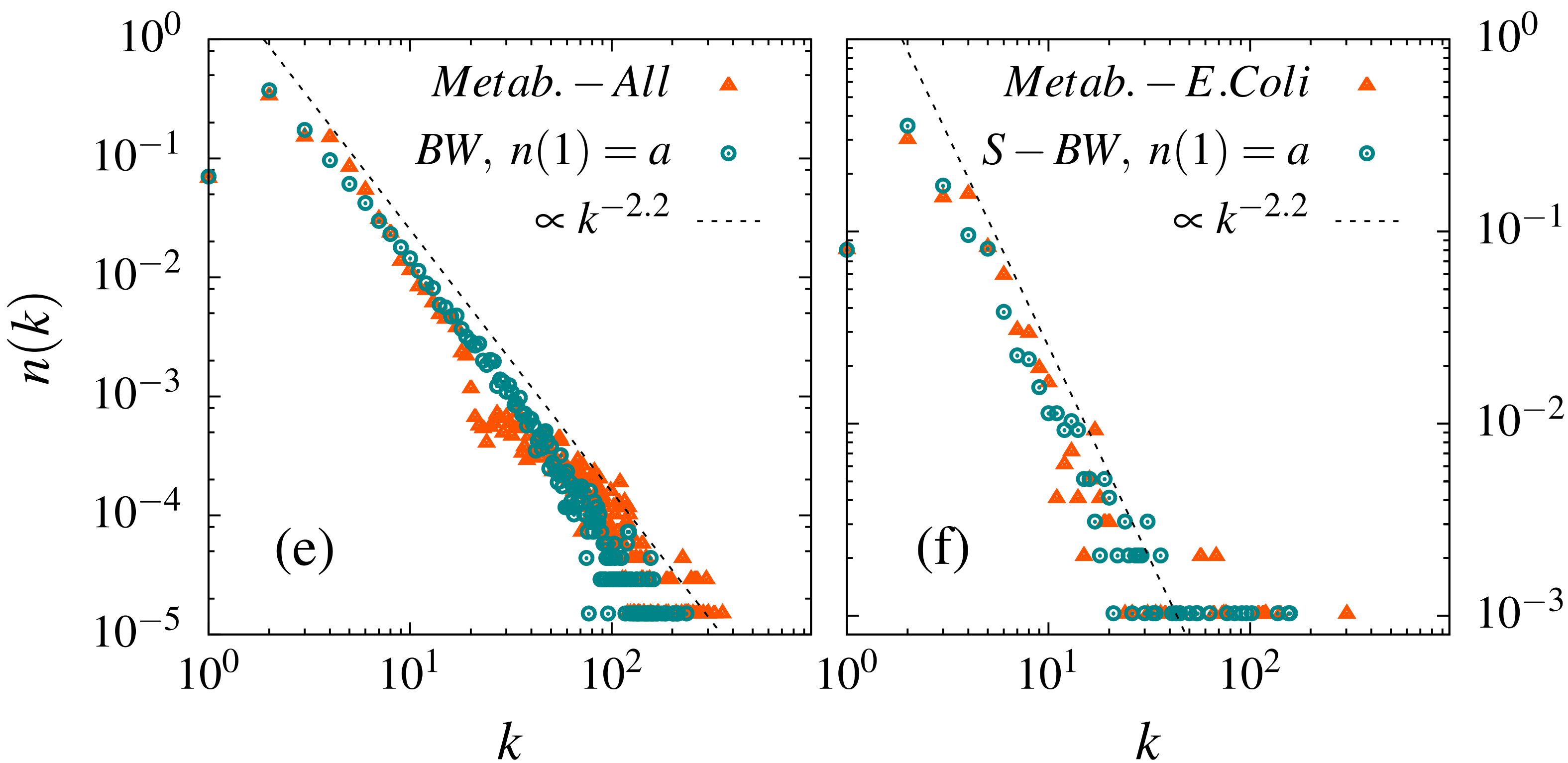


# Comparison with real data

Comparison, with extra constraint:

(e) Same as in c) but with fixed n(1)

(f) Same as in d) but with fixed n(1)



# Conclusions - Part I

PLoS ONE 3(2): e1690,(2008).

# Conclusions - Part I

- Agreement between the BW network and metabolic networks looks very good $\Rightarrow$
  Natural selection has had small effect on the Metabolic networks degree distribution.

PLoS ONE 3(2): e1690,(2008).

# Conclusions - Part I

- Agreement between the BW network and metabolic networks looks very good $\Rightarrow$
  Natural selection has had small effect on the Metabolic networks degree distribution.

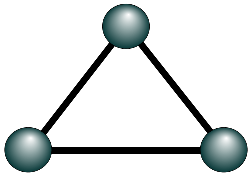
- BW is a random network just as ER is a random network.

PLoS ONE 3(2): e1690,(2008).

# Network properties

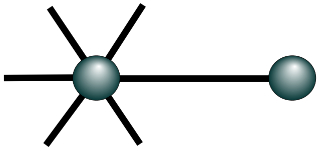
# Network properties

## Clustering Coefficient



$$0 \leqslant C \leqslant 1$$

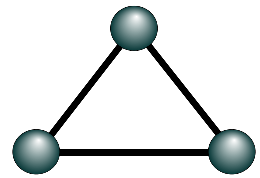
## Assortativity



$$-1 \leqslant r \leqslant 1$$

# Network properties

## Clustering Coefficient

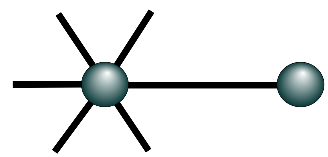


$$0 \leqslant C \leqslant 1$$

$$\langle C \rangle_{metab} = 0.139\ (0.143)$$

$$\langle C \rangle_{BW} = 0.103\ (0.096)$$

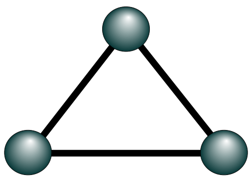
## Assortativity



$$-1 \leqslant r \leqslant 1$$

# Network properties

## Clustering Coefficient

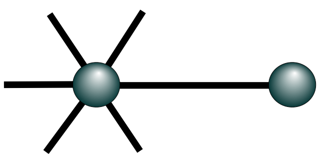


$$0 \leqslant C \leqslant 1$$

$$\langle C \rangle_{metab} = 0.139\ (0.143)$$

$$\langle C \rangle_{BW} = 0.103\ (0.096)$$

## Assortativity

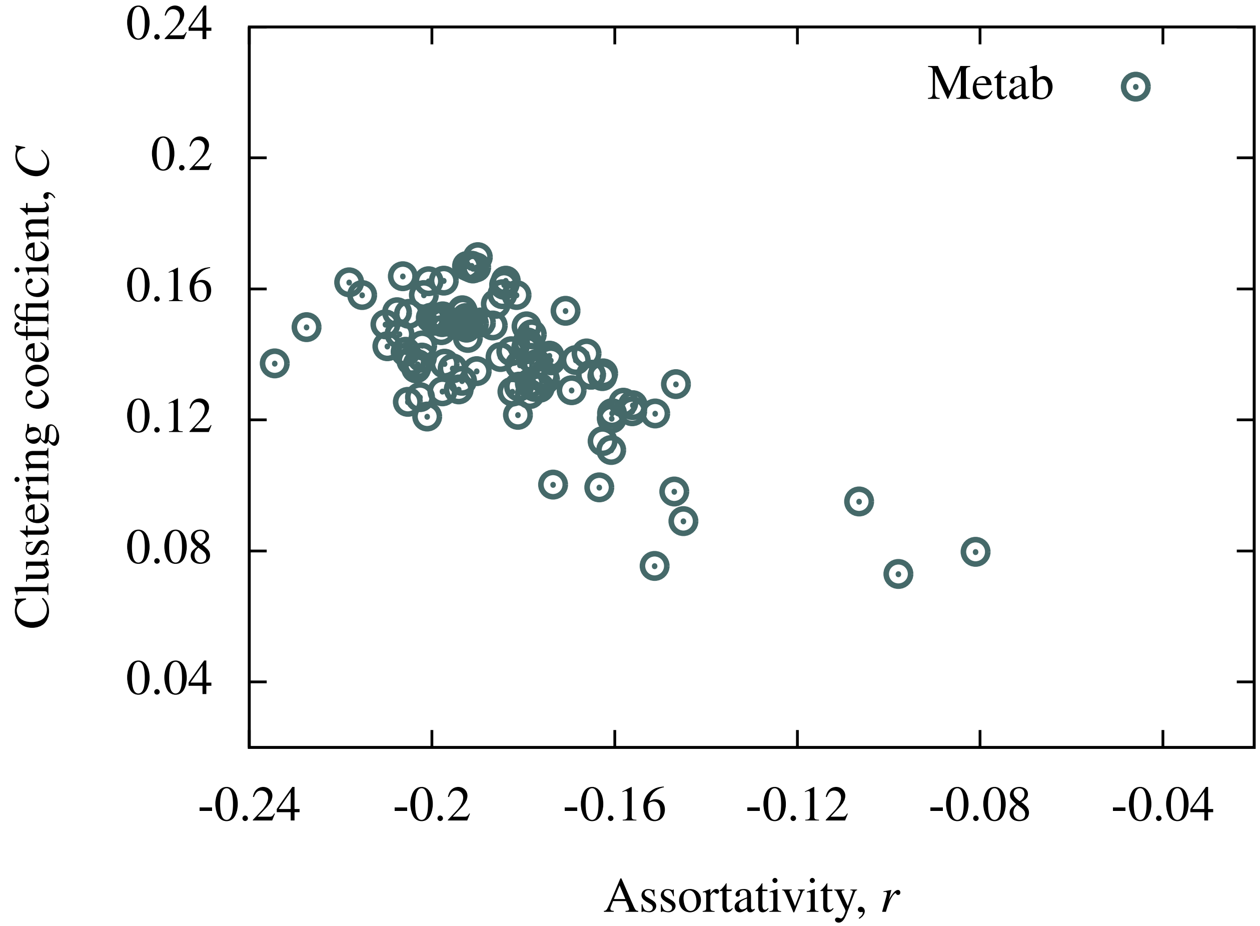


$$-1 \leqslant r \leqslant 1$$

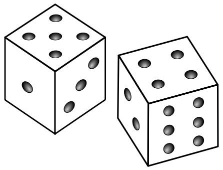
$$\langle r \rangle_{metab} = -0.18\ (-0.178)$$

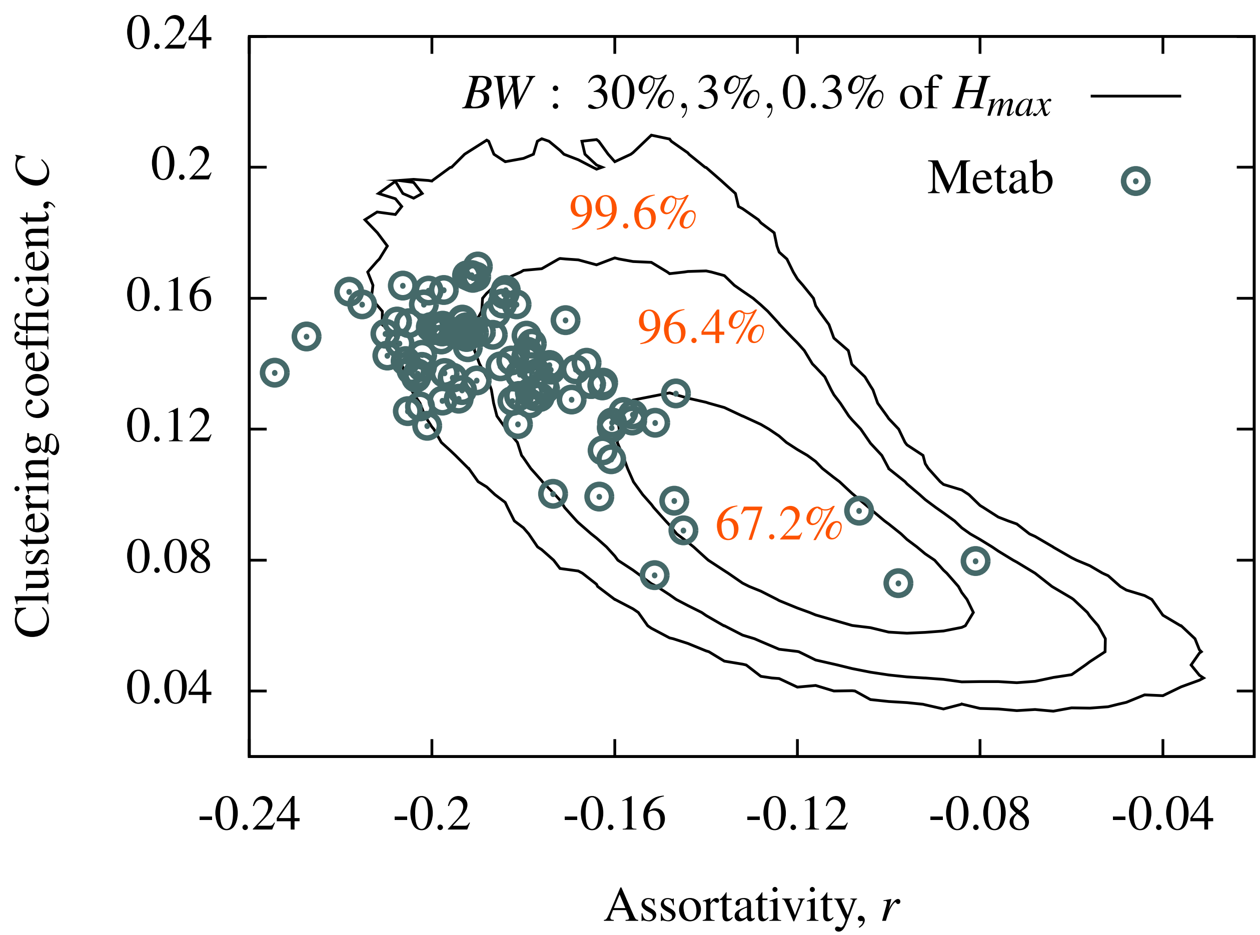
$$\langle r \rangle_{BW} = -0.123\ (-0.125)$$

# Clustering-Assortativity space

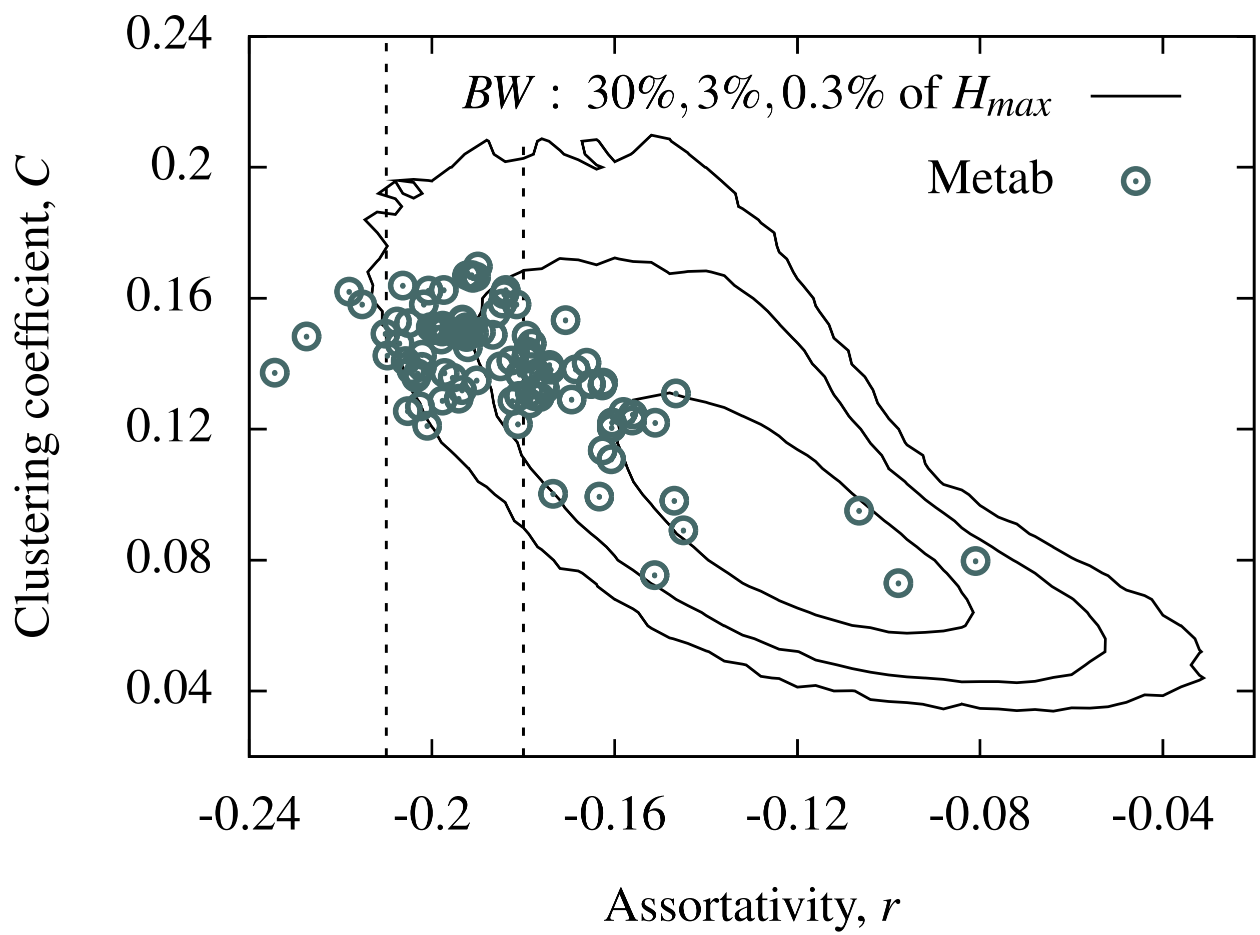


# Clustering-Assortativity space





# Region of Low Assortativity

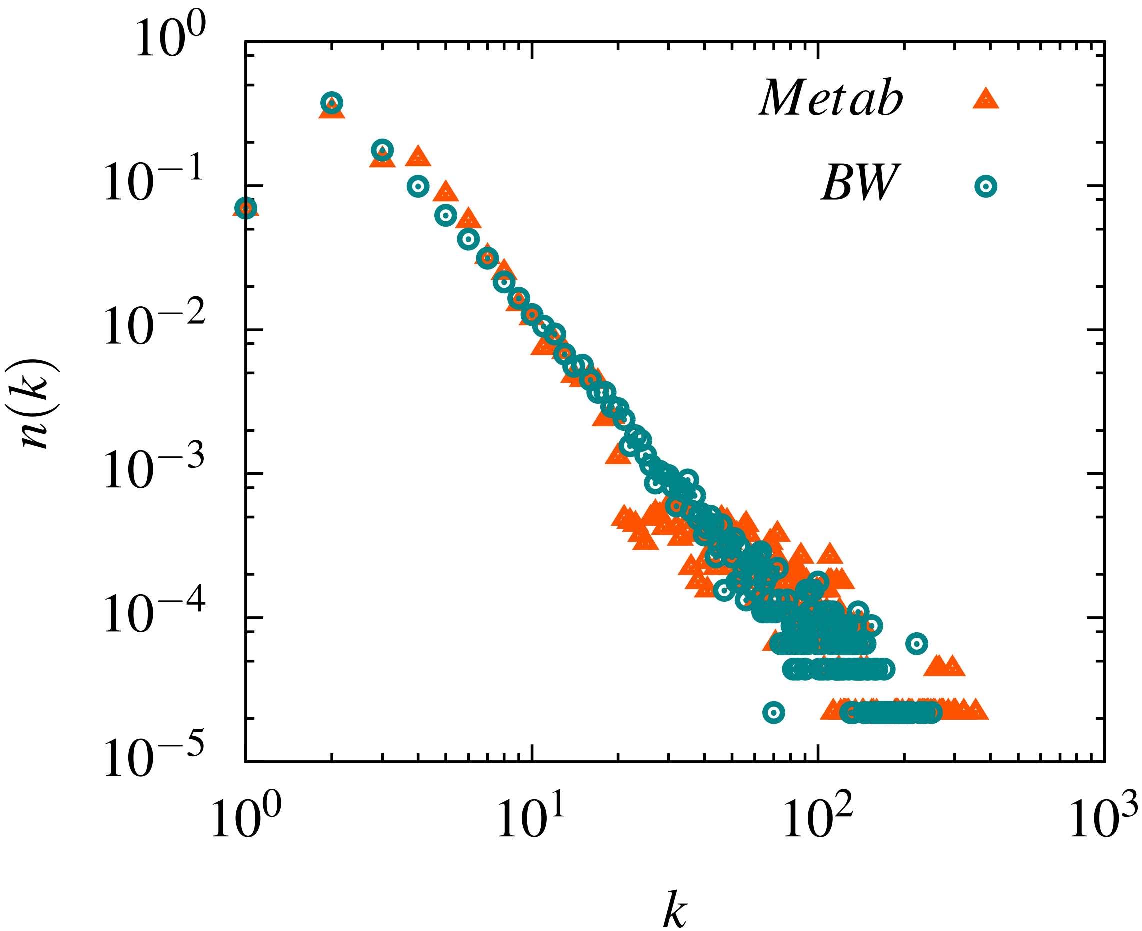



$$-0.21 < r < -0.18 \Rightarrow N_{metab} = 62$$

# Region of Low Assortativity

$\langle C \rangle_{metab} = 0.148$

$\langle C \rangle_{BW} = 0.149$



# Conclusions - Part II

# Conclusions - Part II

- Metabolic networks are close to the null model (BW).

# Conclusions - Part II

- Metabolic networks are close to the null model (BW).
- Deviations indicates evolutionary pressure towards lower assortativity.

# Conclusions - Part II

- Metabolic networks are close to the null model (BW).
- Deviations indicates evolutionary pressure towards lower assortativity.
- This pressure, to large extent, is reflected in a small change in the degree distribution.

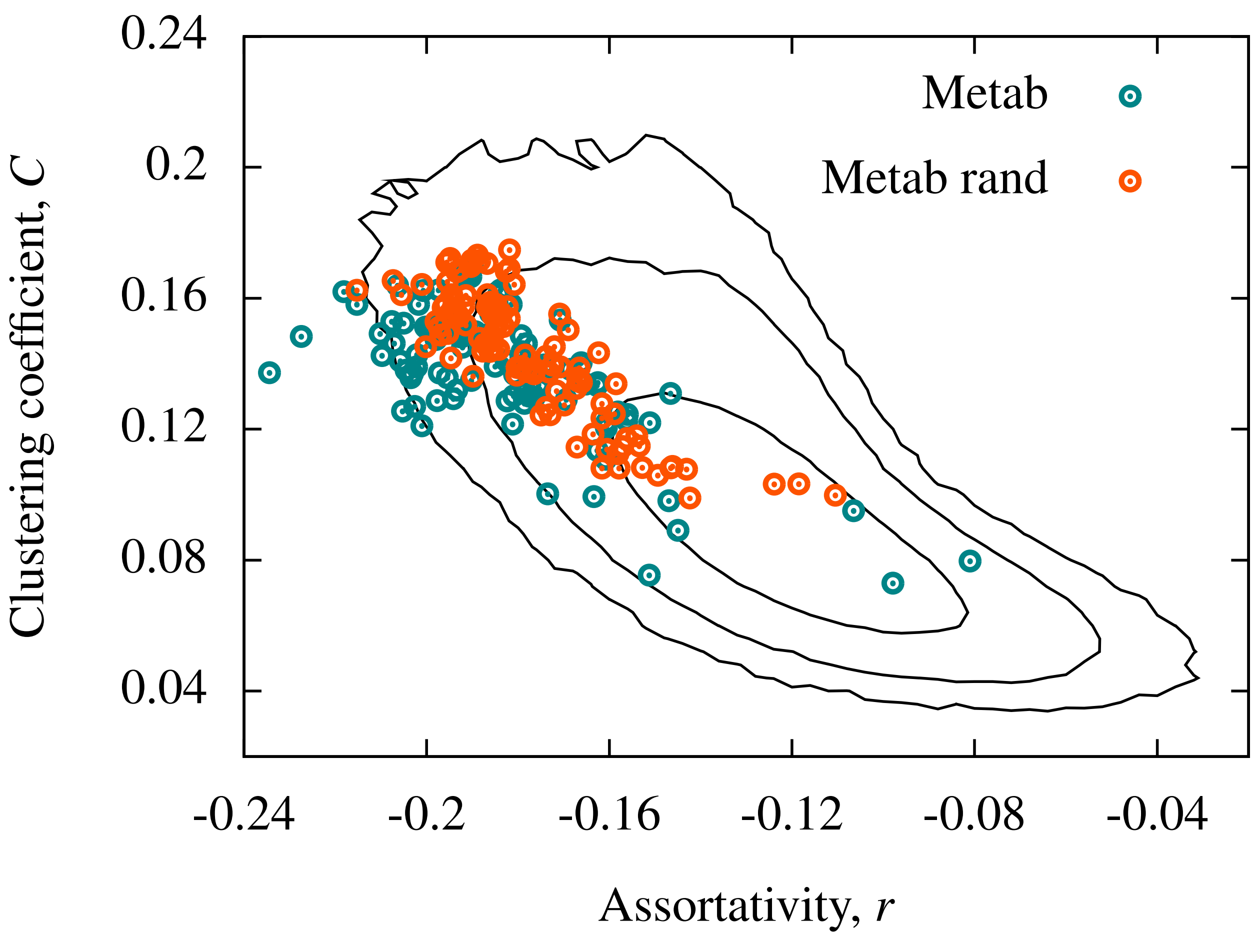
# Conclusions - Part II

- Metabolic networks are close to the null model (BW).
- Deviations indicates evolutionary pressure towards lower assortativity.
- This pressure, to large extent, is reflected in a small change in the degree distribution.
- No evolutionary pressure on clustering.

# Acknowledgment

I would like to thank...

- The organizers for inviting me.
- Petter Minnhagen for the collaboration on this work.
- Luis Rocha and Seung Ki Baek for comments and suggestions.


Metab
Metab rand
Clustering coefficient, C
Assortativity, r
0.24
0.2
0.16
0.12
0.08
0.04
-0.24
-0.2
-0.16
-0.12
-0.08
-0.04