Reconstruction and Analysis of Metabolic Networks using Graph Theory

Nara C. Marella

Advisor: Suguna Challa

Work performed in the Laboratory of Prof. Somdatta Sinha (CCMB)

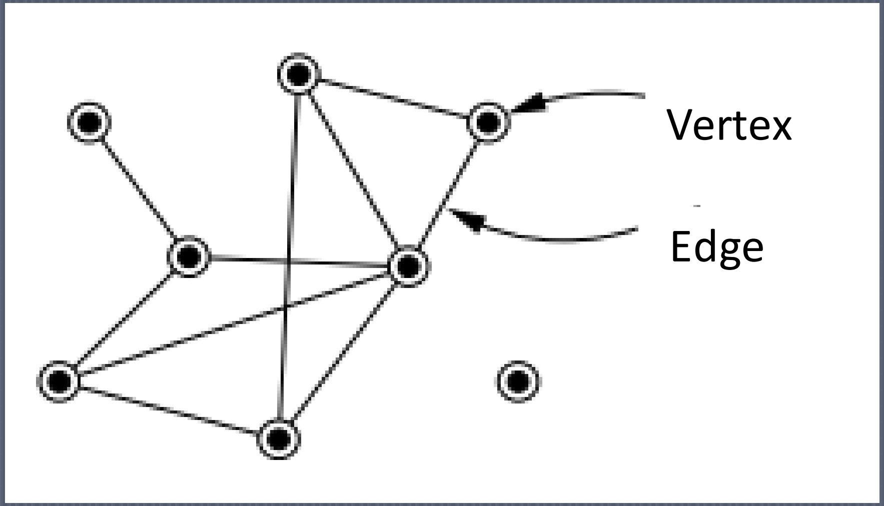
24. May 2010

Outline

- Background
  - Types of Networks
  - How to Construct Metabolic Network ?
  - Network Properties
- Statement Of Problem
- Results
- Discussion

What is a Network

A network is a set of items, which we call vertices (or sometimes nodes) with connections between them, called edges



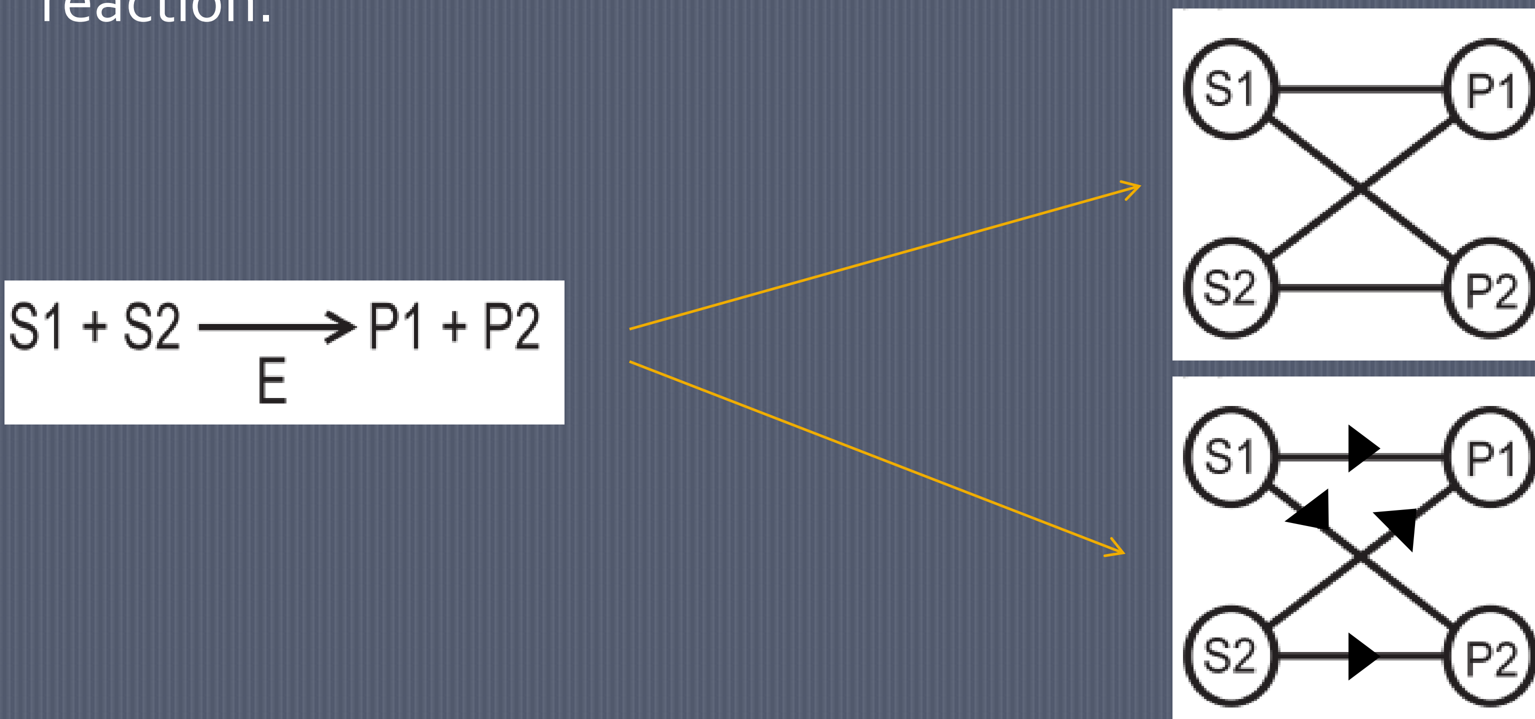
Networks – Some Examples

1. Social Networks	Patterns of friendships Social Networking on WWW
2. Information Networks	Network of Citations
3. Technological Networks	Electric Power Grids Network of Airline Routes Network of Roads, Railways etc
4. Biological Networks	Metabolic Networks Gene Regulatory Networks Ecological Networks (Food Web) Protein Contact Networks

Representing biochemical reaction in the form of Network

Metabolite → node

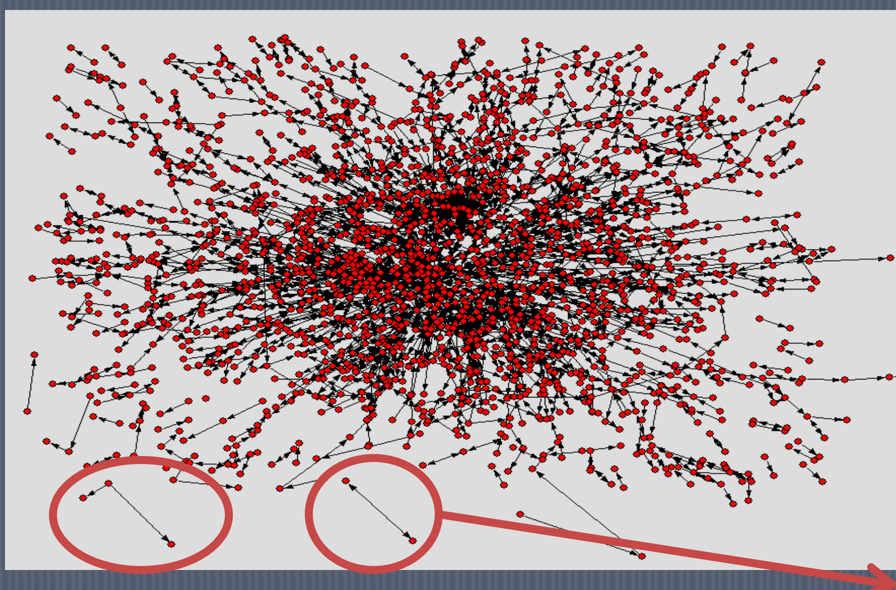
Nodes are connected if they form **substrate-product pair** in a reaction.



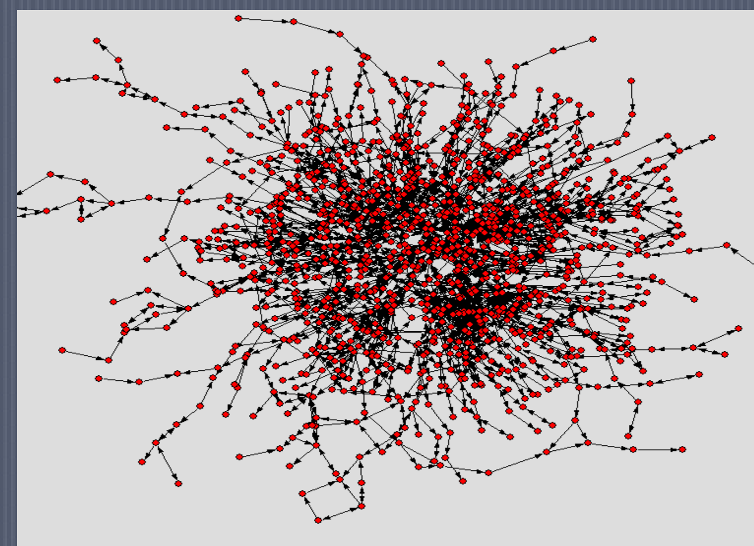
Parameters to study structure of Networks

Giant Strong Component

E.coli complete metabolic network



E.coli giant strong component



Isolated components

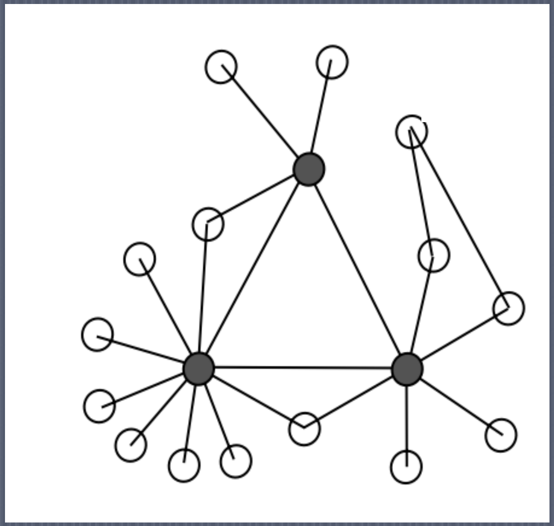
It is the largest fraction of the network within which communication is possible

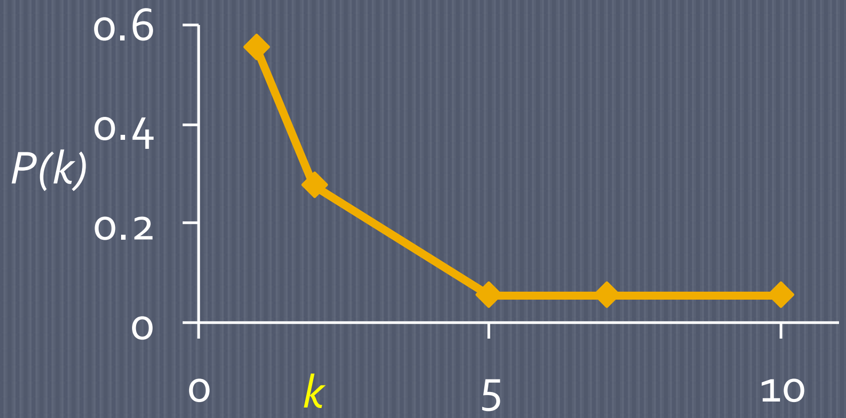
Parameters to study structure of Networks

Degree Distributions

$$P(k) = n_k / n$$

k = degree
n = total no of nodes
n_k = no of nodes of degree k
$P(k)$ = fraction of vertices with degree k





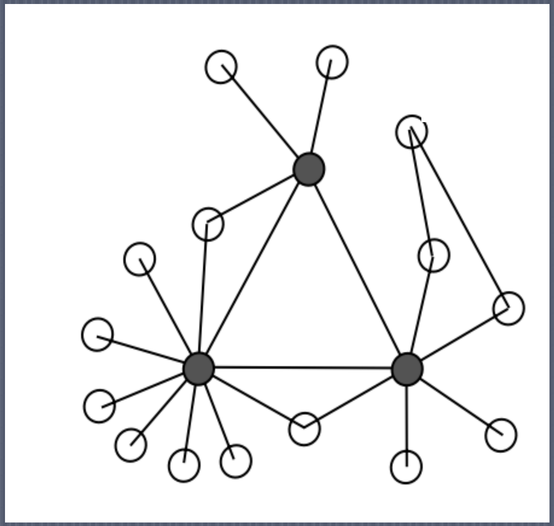
Parameters to study structure of Networks

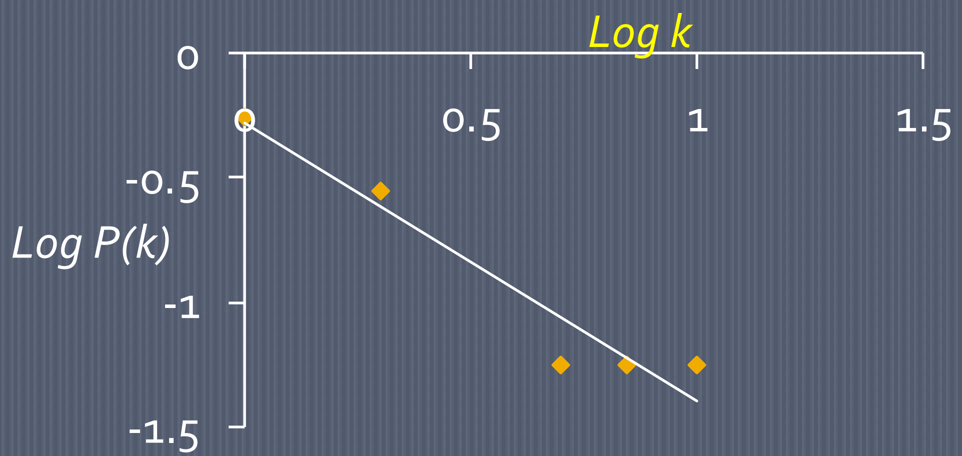
Degree Exponent

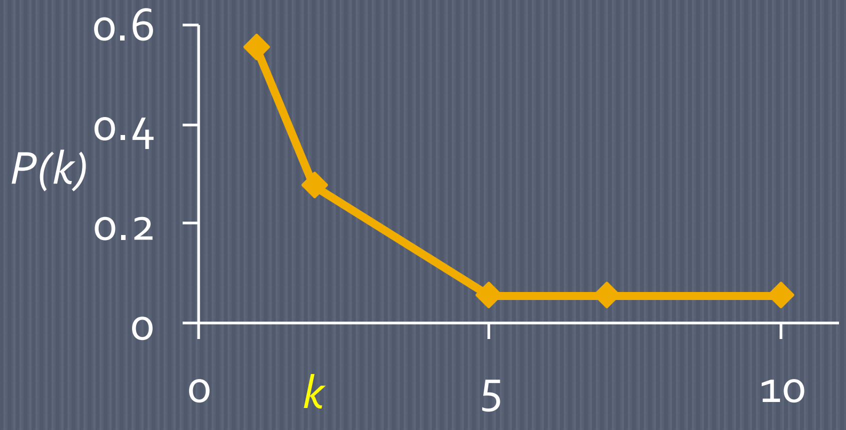
$$P(k) \sim k^{-\gamma}$$

γ - Degree Exponent

High DE → heterogeneity







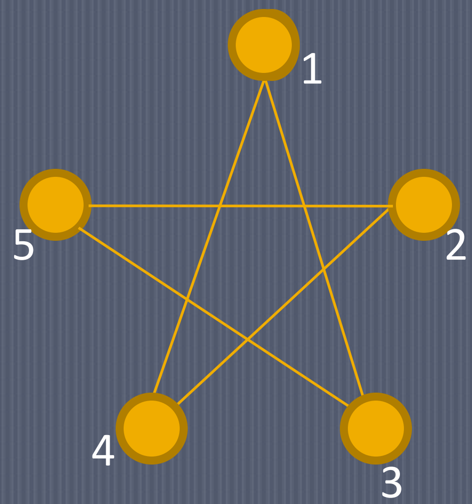
Parameters to study structure of Networks

<u>Significance of scale freeness</u>

1. Random node disruptions do not lead to a major loss of connectivity
2. Metabolic networks are scale free → robust in face of random disruptions
3. Loss of the high degree nodes causes the breakdown of the network into isolated clusters

Parameters to study structure of Networks

Shortest path



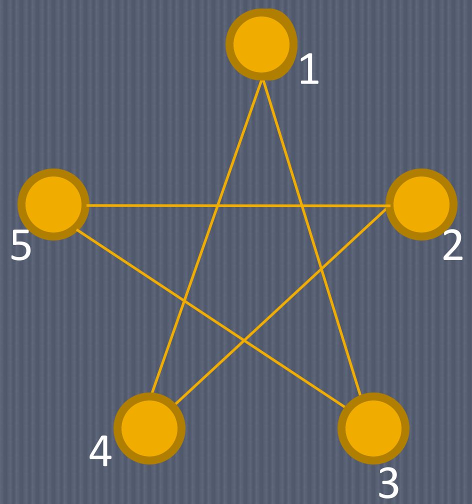
The minimum number of steps required to reach a node from another

Pairs	Shortest Path	
1->2	1->4->2	2 steps
1->3	1->3	1 step
1->4	1->4	1 step
1->5	1->3->5	2 steps
2->3	2->5->3	2 steps
2->4	2->4	1 step
2->5	2->5	1 step
3->4	3->1->4	2 steps
3->5	3->5	1 step
4->5	4->2->5	2 steps

Parameters to study structure of Networks

Average Path Length 1.5

Average of shortest paths for all pairs of nodes



It is a measure of the efficiency of information transfer in a network

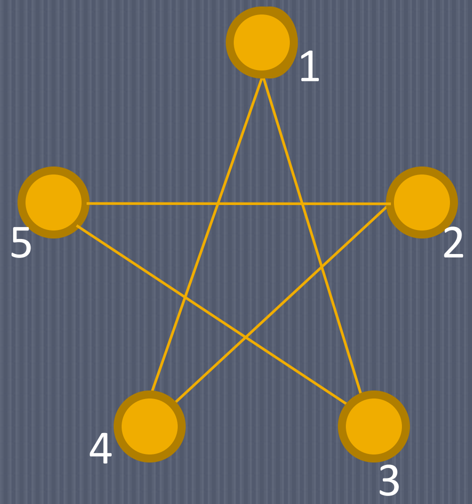
Pairs	Shortest Path	
1->2	1->4->2	2 steps
1->3	1->3	1 step
1->4	1->4	1 step
1->5	1->3->5	2 steps
2->3	2->5->3	2 steps
2->4	2->4	1 step
2->5	2->5	1 step
3->4	3->1->4	2 steps
3->5	3->5	1 step
4->5	4->2->5	2 steps

Parameters to study structure of Networks

Diameter

2

Longest of the shortest path in a network



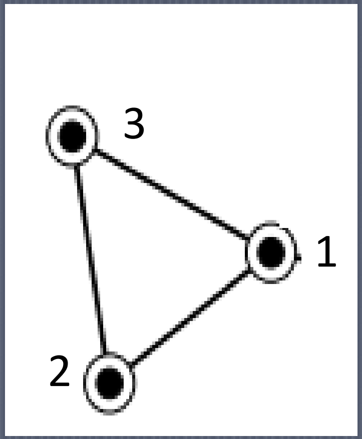
Pairs	Shortest Path	
1->2	1->4->2	2 steps
1->3	1->3	1 step
1->4	1->4	1 step
1->5	1->3->5	2 steps
2->3	2->5->3	2 steps
2->4	2->4	1 step
2->5	2->5	1 step
3->4	3->1->4	2 steps
3->5	3->5	1 step
4->5	4->2->5	2 steps

Parameters to study structure of Networks

Transitivity

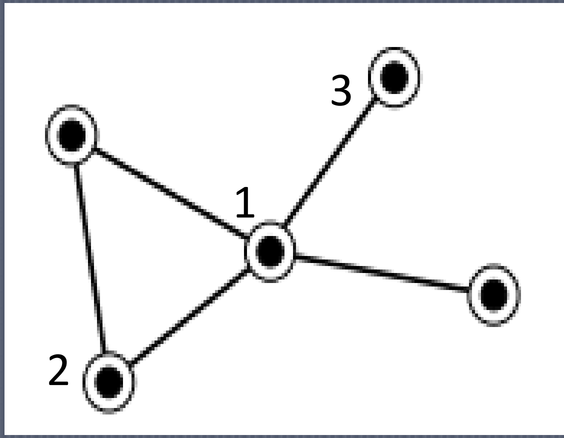
1 --- 2
1 --- 3

2 --- 3



Parameters to study structure of Networks

Clustering or Transitivity



Clustering coefficient of a node = $\frac{\text{Number of existing connections between it's neighbors}}{\text{Number of possible connections between it's neighbors}}$

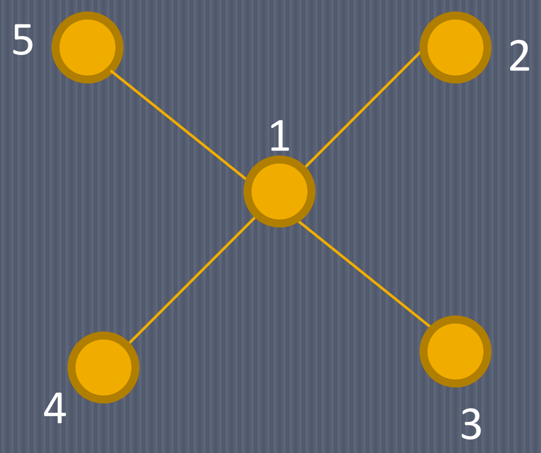
Clustering Coefficient of Node 1 is (1/ 6) , Node 2 is (1 / 1) = 1

Clustering Coefficient of network = 0.433 (Average of 1, 1, 1/6, 0 and 0)

Parameters to study structure of Networks

Betweenness Centrality

Number of shortest paths that run through a vertex



Shortest Paths passing through '1'

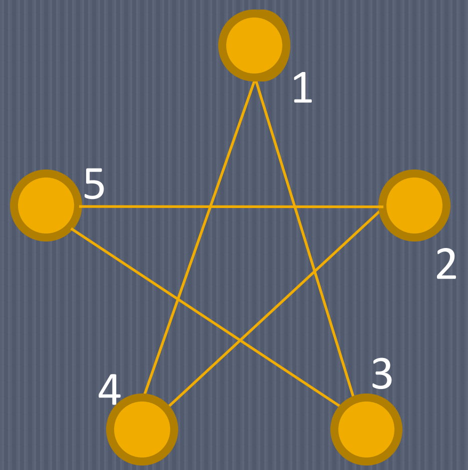
a. 2 – 3
b. 2 – 4
c. 2 – 5
d. 3 – 4
e. 3 – 5
f. 4 – 5

Betweenness centrality of Node 1 is (6/6) = 1

Parameters to study structure of Networks

Closeness centrality

Mean shortest distance between a vertex and all other vertices reachable from it



1->2 shortest – 1->4->2 – 2 steps

1->3 shortest – 1->3 – 1 steps

1->4 shortest – 1->4 – 1 steps

1->5 shortest – 1->3->5 – 2 steps

Closeness centrality of node 1 = (2+1+1+2)/4 = 1.5

What are factors that shape structure of metabolic networks ?

Comparative genomics of metabolic networks of free-living and parasitic eukaryotes

Barbara Nerima, Daniel Nilsson, Pascal Maser
(University of Bern, Switzerland)

BMC *Genomics 2010, 11:217*

"Parasites networks were smaller than those of non-parasites regarding number of nodes or edges"

"Life style of an organism might be playing a role in shaping the structure of metabolic network"

What are factors that shape structure of metabolic networks ?

Correlation between structure and temperature in prokaryotic metabolic networks

Kazuhiro Takemoto, Jose C Nacher and Tatsuya Akutsu
(Kyoto University)

BMC *Bioinformatics* 2007, 8:303

Does temperature at which an organism grow play a role in shaping the structure of metabolic network ?

GOLD

Genomes OnLine Database

Resource for information on genome sequencing projects

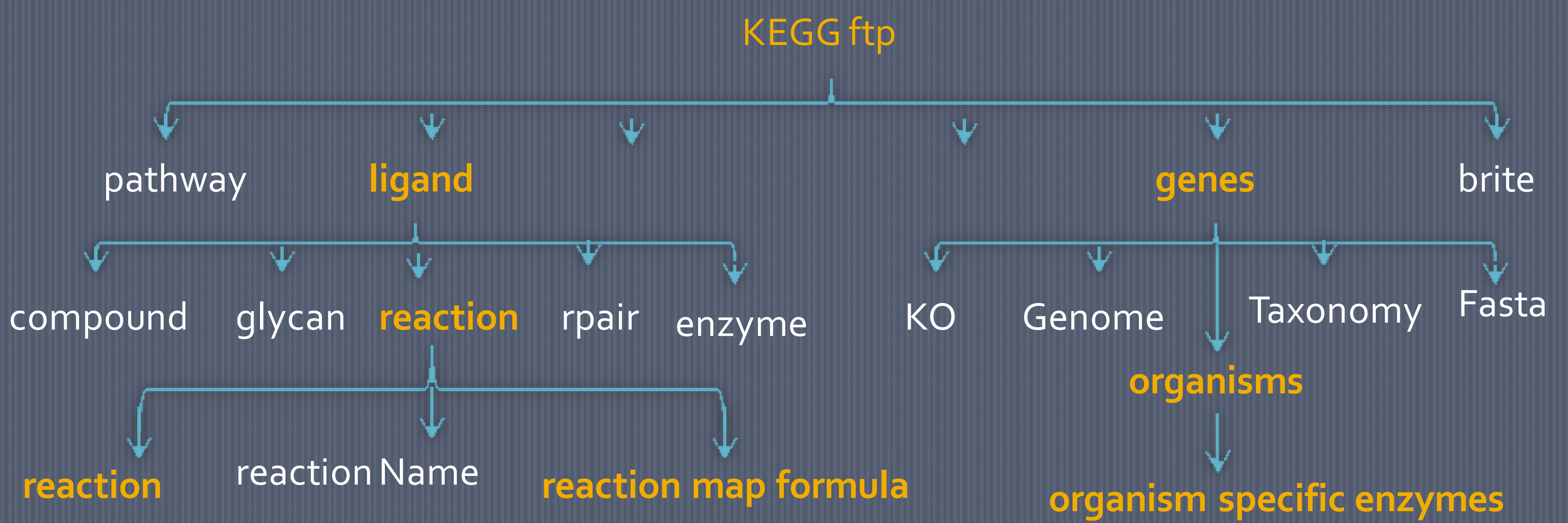
Temperature of Organism is taken from here

- psychrophiles
- mesophiles
- thermophiles
- hyperthermophiles

KEGG

Organisms synthesize enzymes which are required for its survival

Directories in KEGG database



Reconstructing metabolic network

$$S1 + S2 \xrightarrow{E} P1 + P2$$

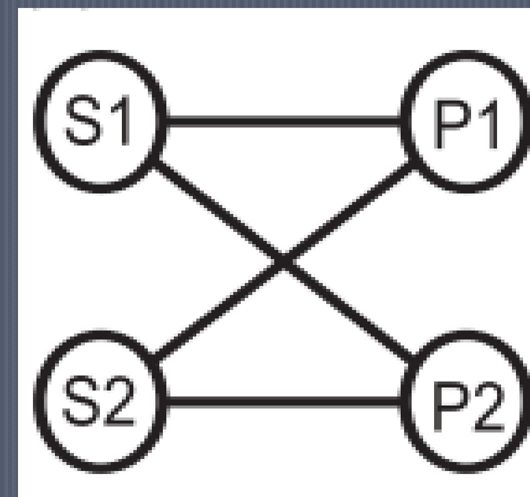
KEGG

→ Perl Script

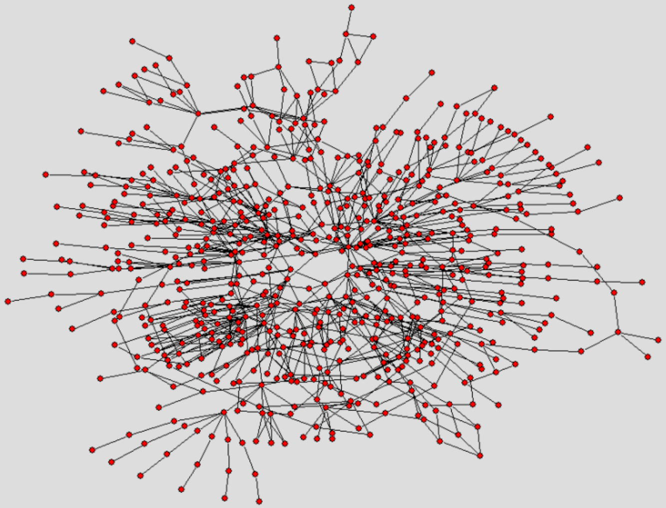
<u>Edge list</u>

S1	P1
S1	P2
S2	P1
S2	P2

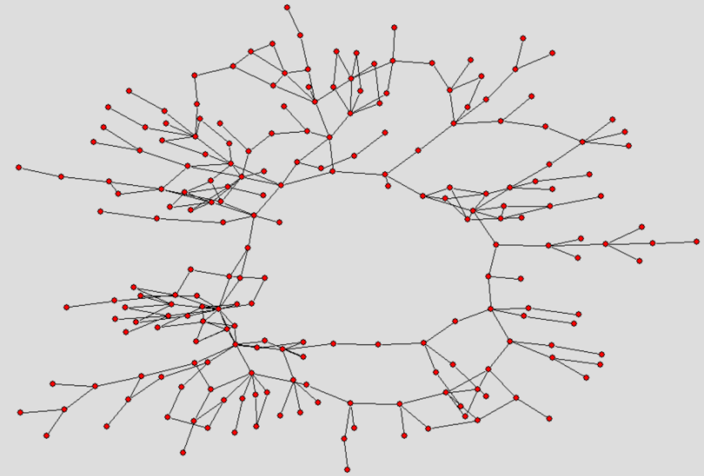
→ Pajek



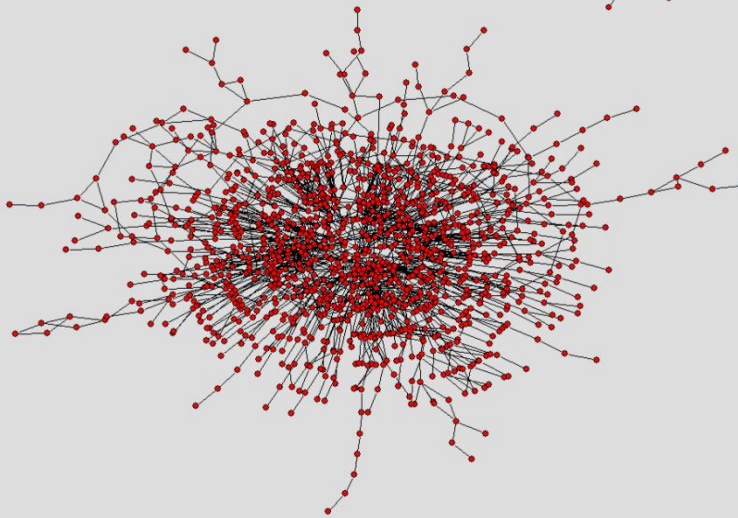
Pajek is a software for visualization and analysis of large networks



Streptococcus pyrogenes

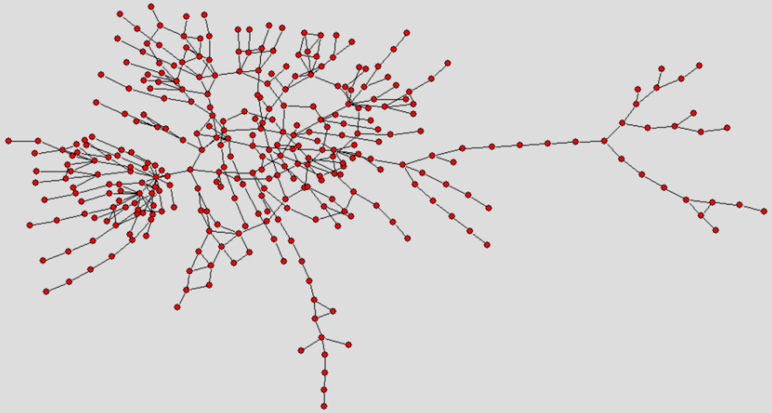


Mycoplasma genitalium

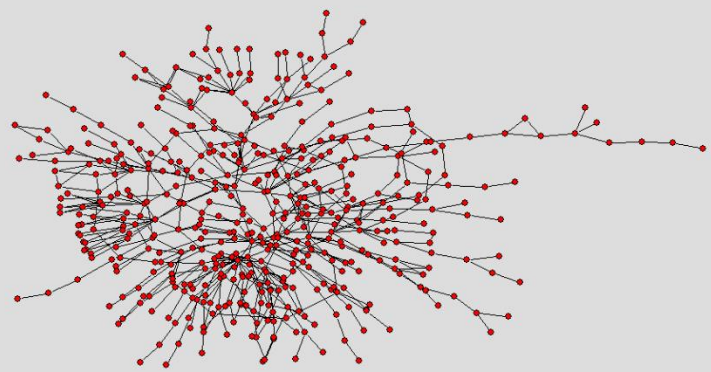


E.coli

Methanopyrus kandleri

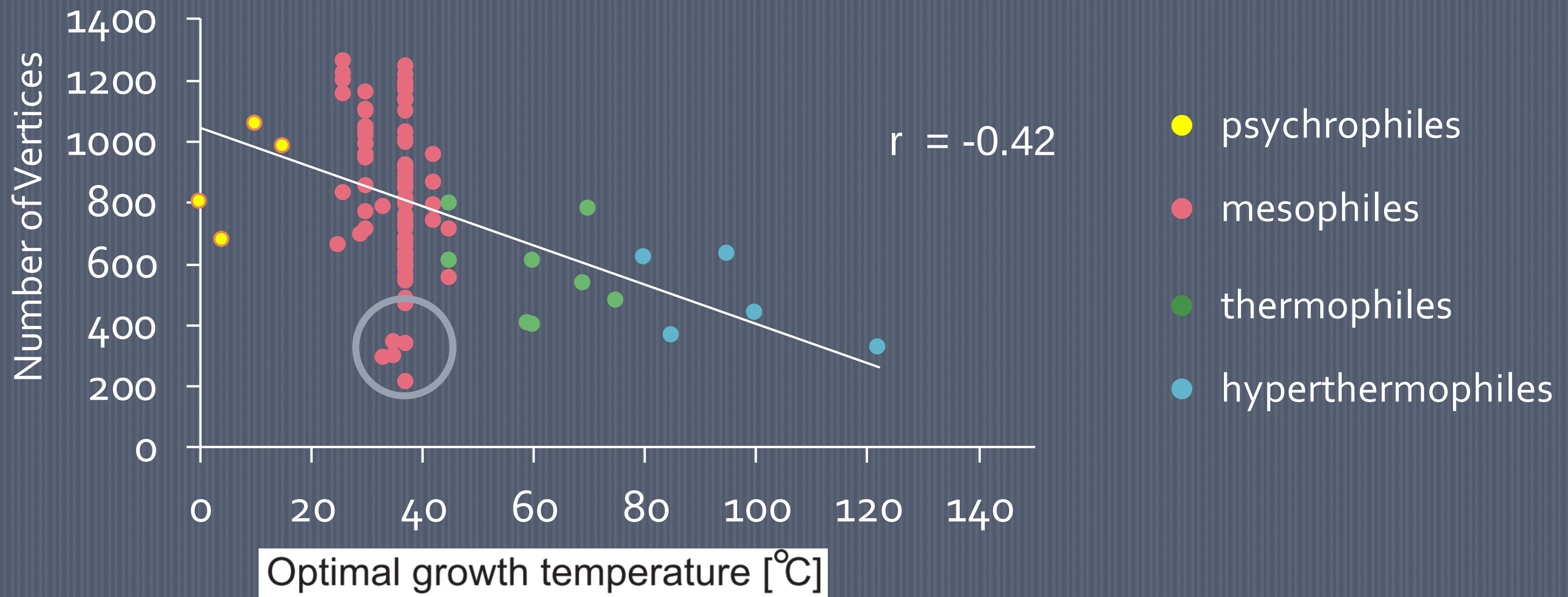


Sulfolobus acidocaldarius



Results

Size of Giant Strong Component Vs Temperature

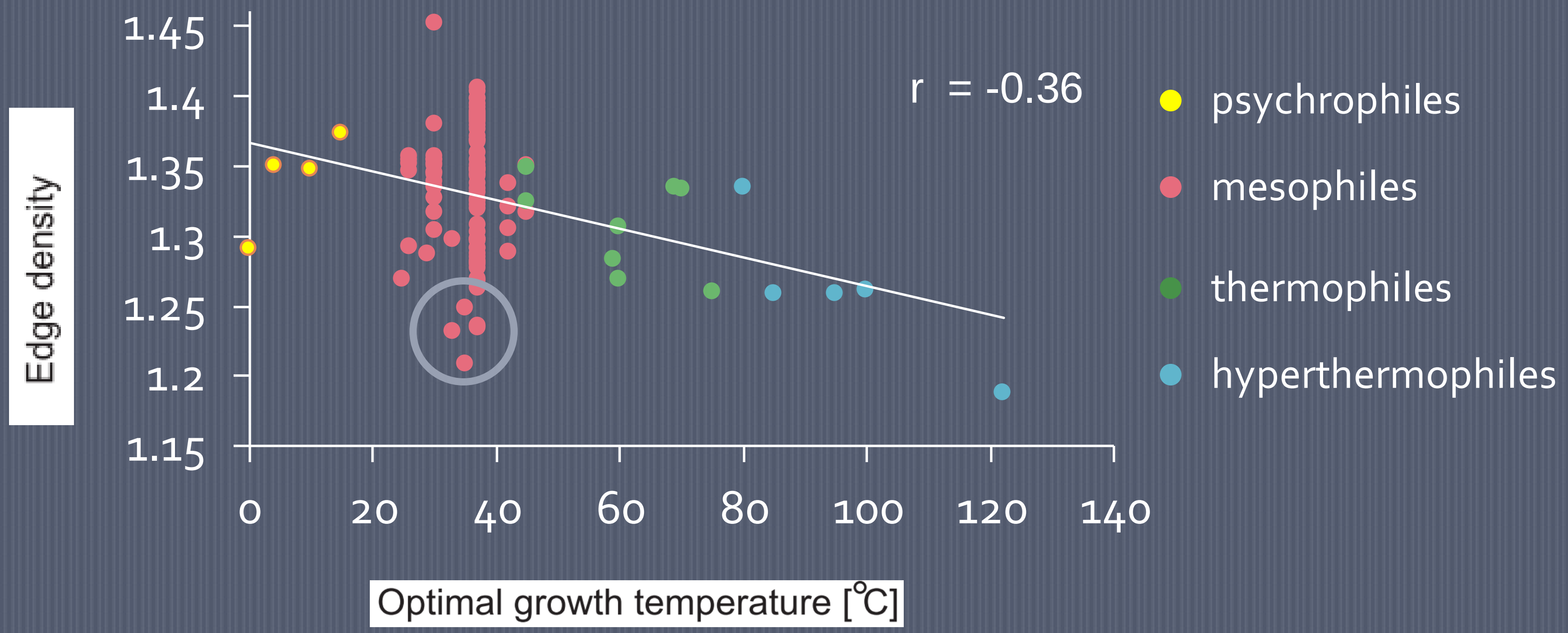


Thermophiles and Hyperthermophiles have less number of metabolites (in g.s.c) than mesophiles and psychrophiles

Reason might be – the environment in which thermophiles and hyperthermophiles live does not have much variation as compared to mesophiles, so thermophiles and hyperthermophiles use less number of metablites to carry out their daily tasks.

Results

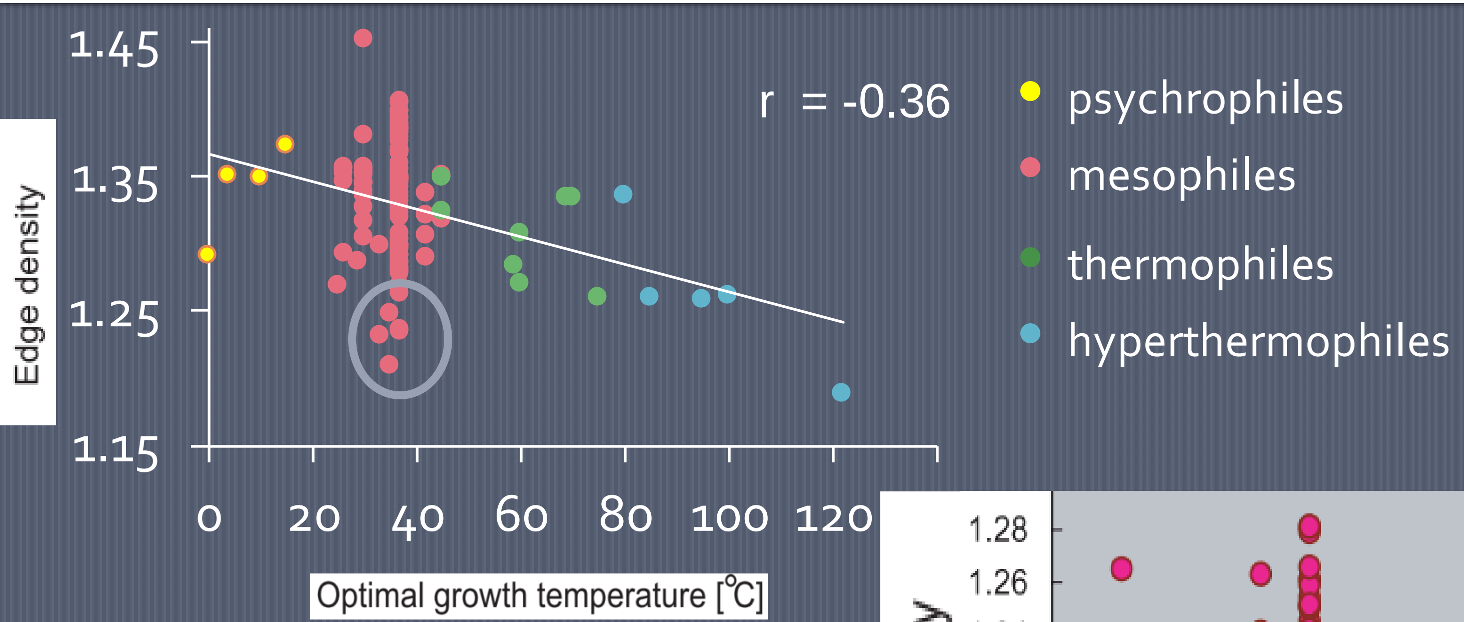
Edge density Vs Temperature

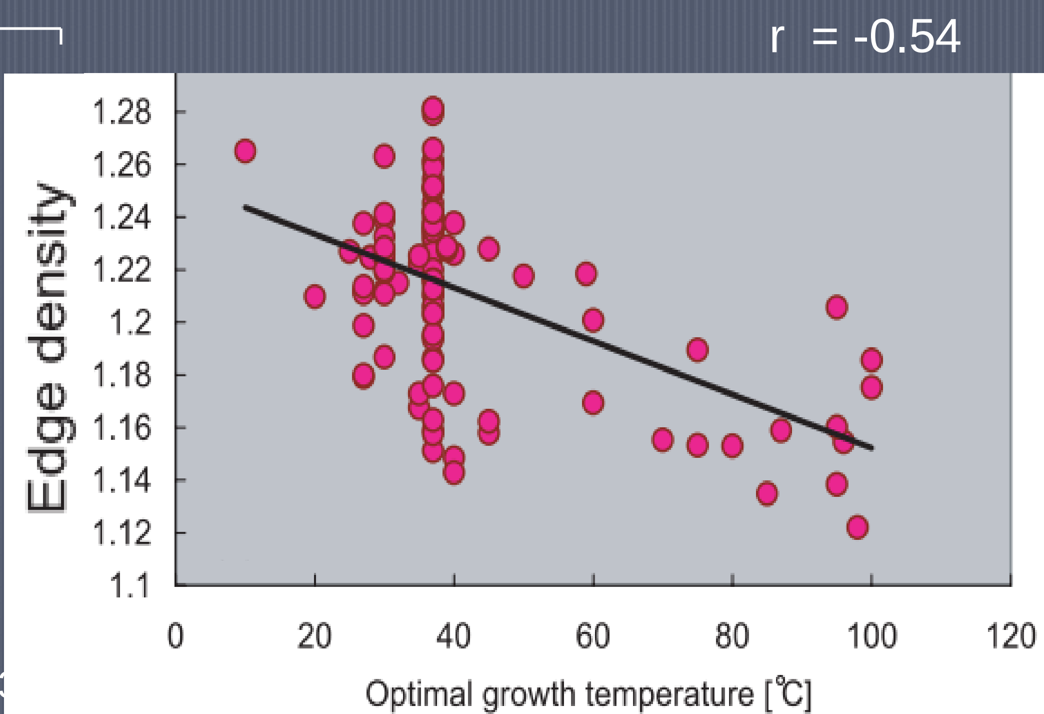


1. **On an average, metabolites participate in fewer reactions as the temperature increases**
2. **There is wide range of variation in mesophiles, there could be similar variability in other groups also if same number of organisms are taken in all groups**

Results

Edge density Vs Temperature

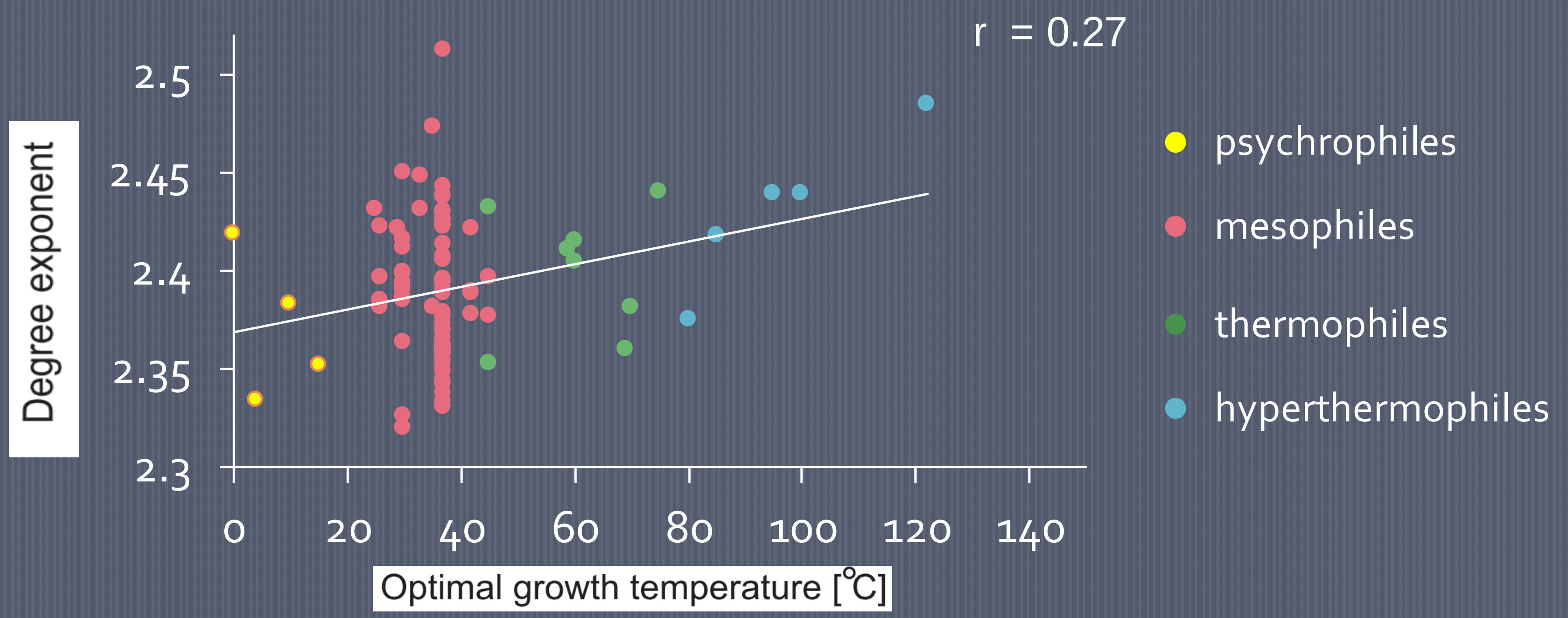




Kazuhiro Takemoto et. al.
BMC *Bioinformatics* 2007, 8:30

Results

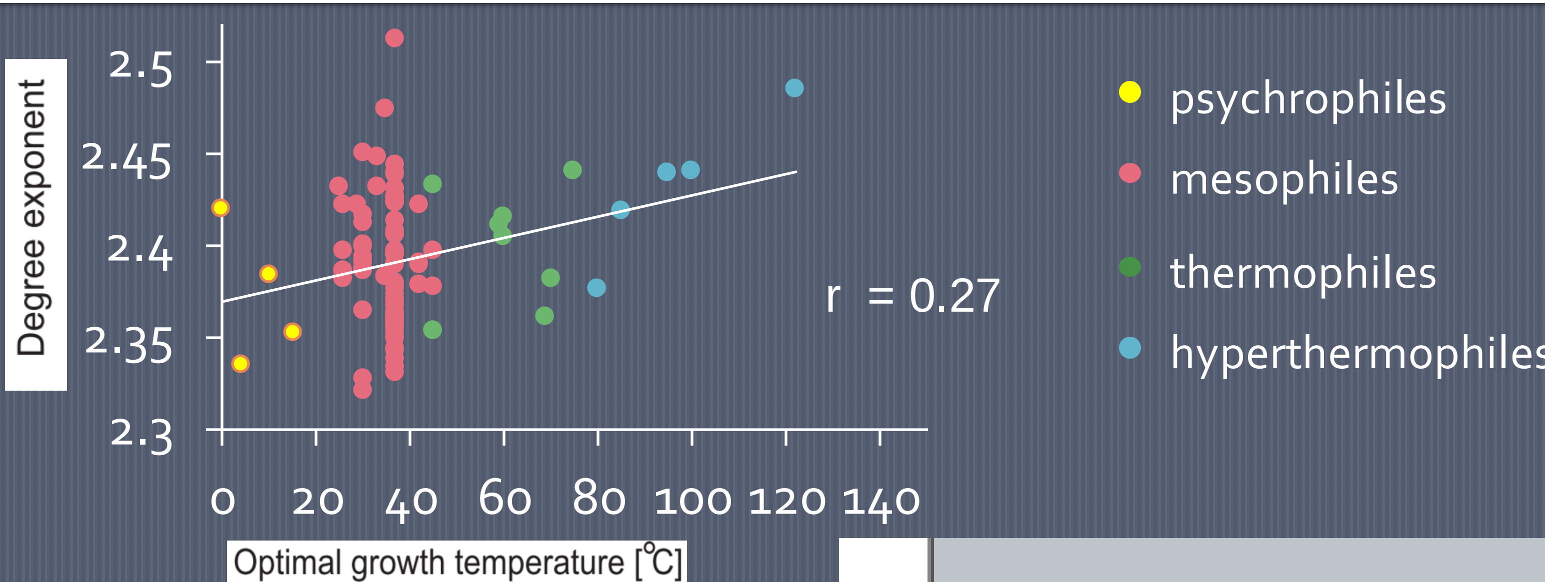
Degree Exponent Vs Temperature

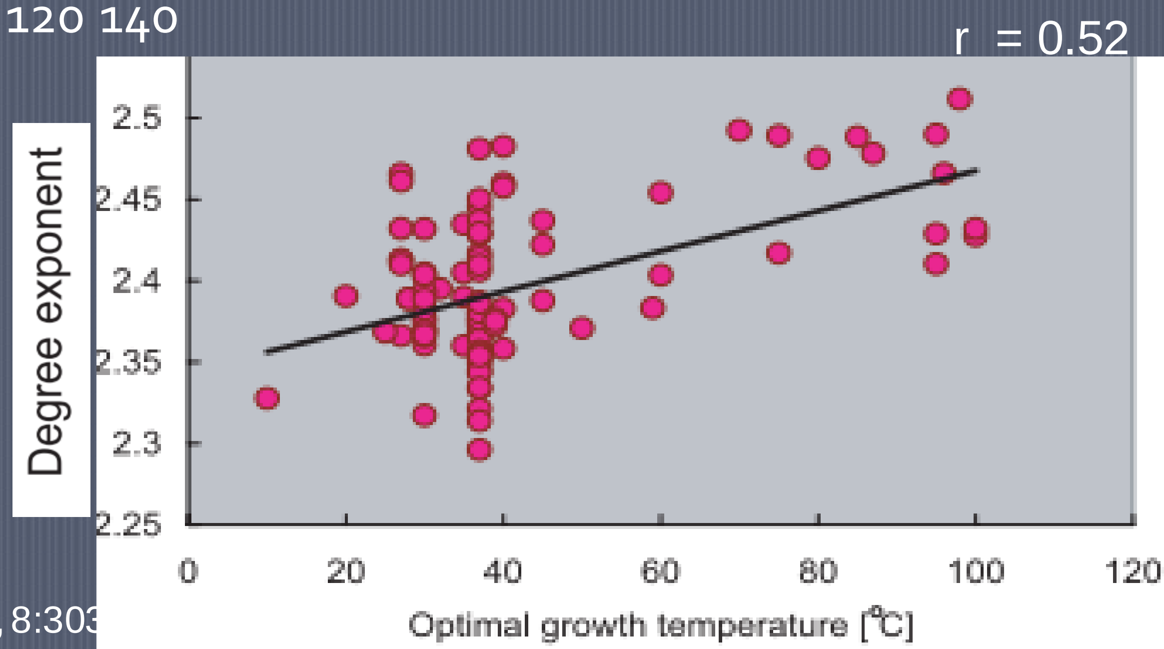


1. There are fewer high degree nodes and more low degree nodes
2. As temp increases, there is less variation in degrees of nodes
3. As temperature increases heterogeneity decreases

Results

Degree Exponent Vs Temperature

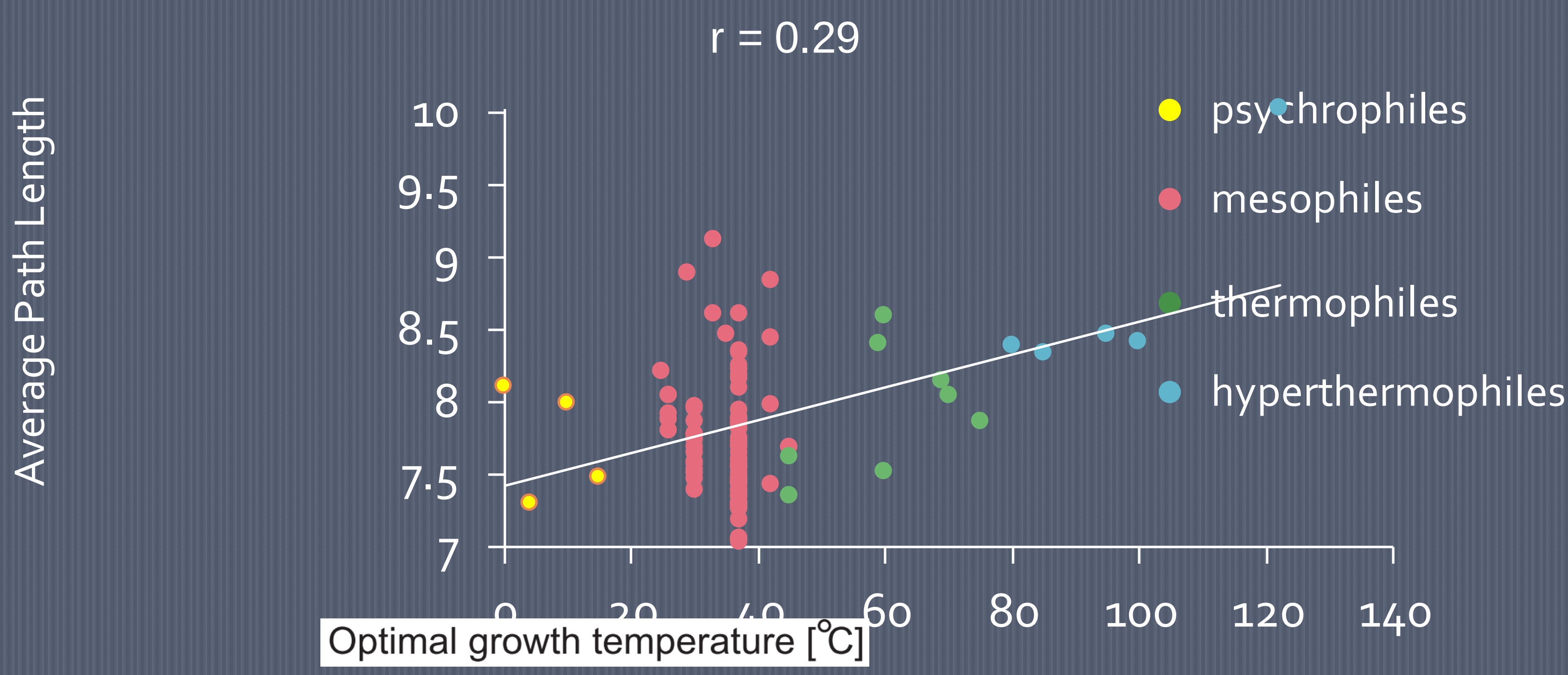




Kazuhiro Takemoto et. al.
BMC *Bioinformatics* 2007, 8:303

Results

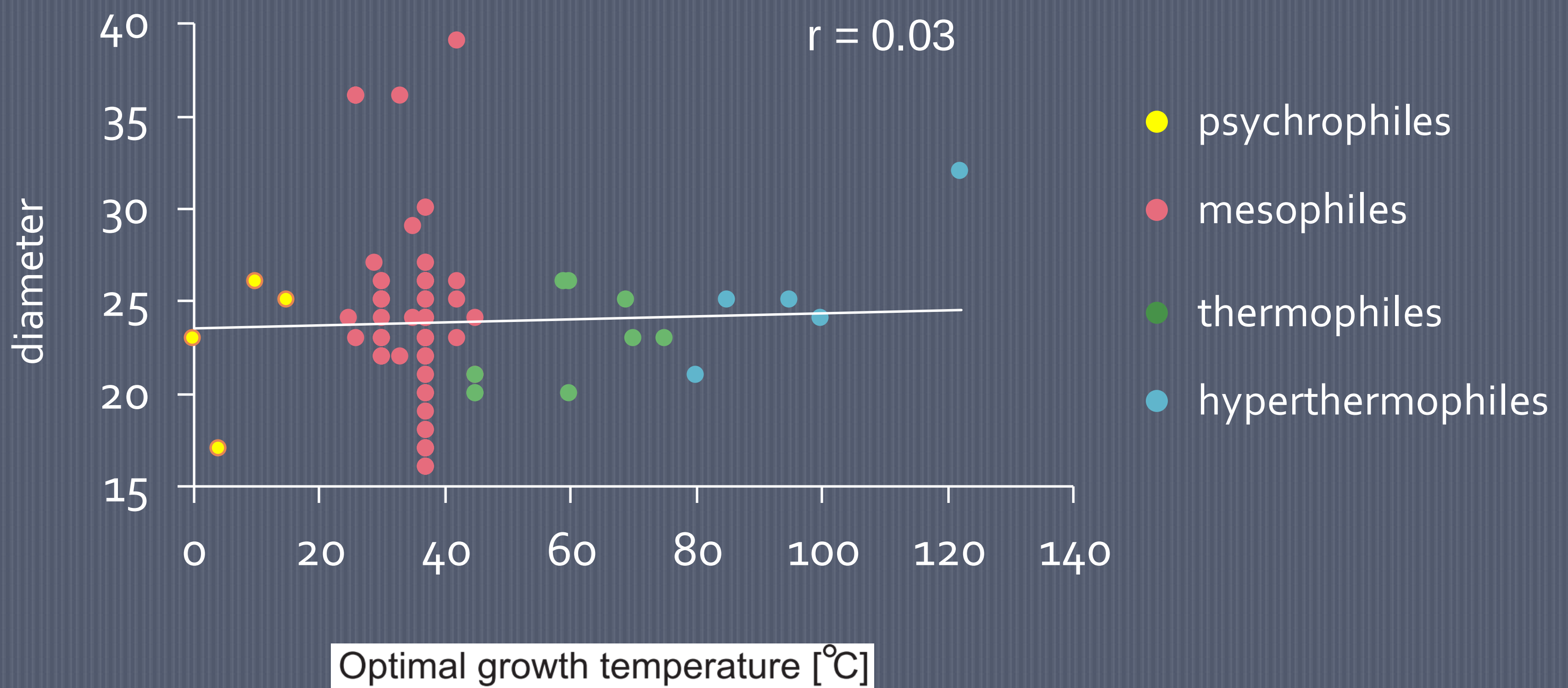
Average Path Length Vs Temperature



1. **Reflection of Evolution**
2. **As temperature increases, number of steps are required for conversion of one metabolite to other increases**

Results

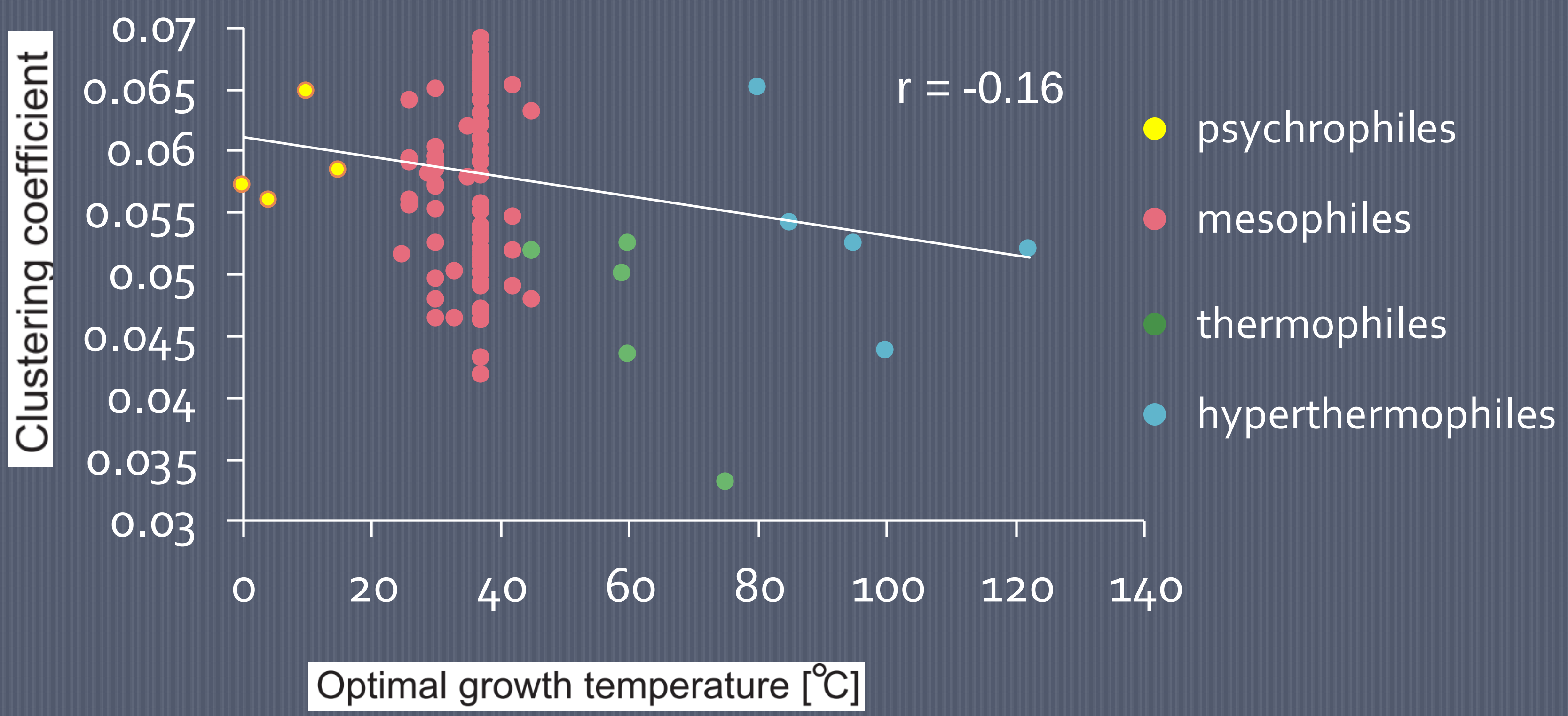
Diameter Vs Temperature



1. **Barabasi et. al - Network Diameter is conserved in all groups**
2. **Large diameter attenuates ability to respond efficiently to external changes**

Results

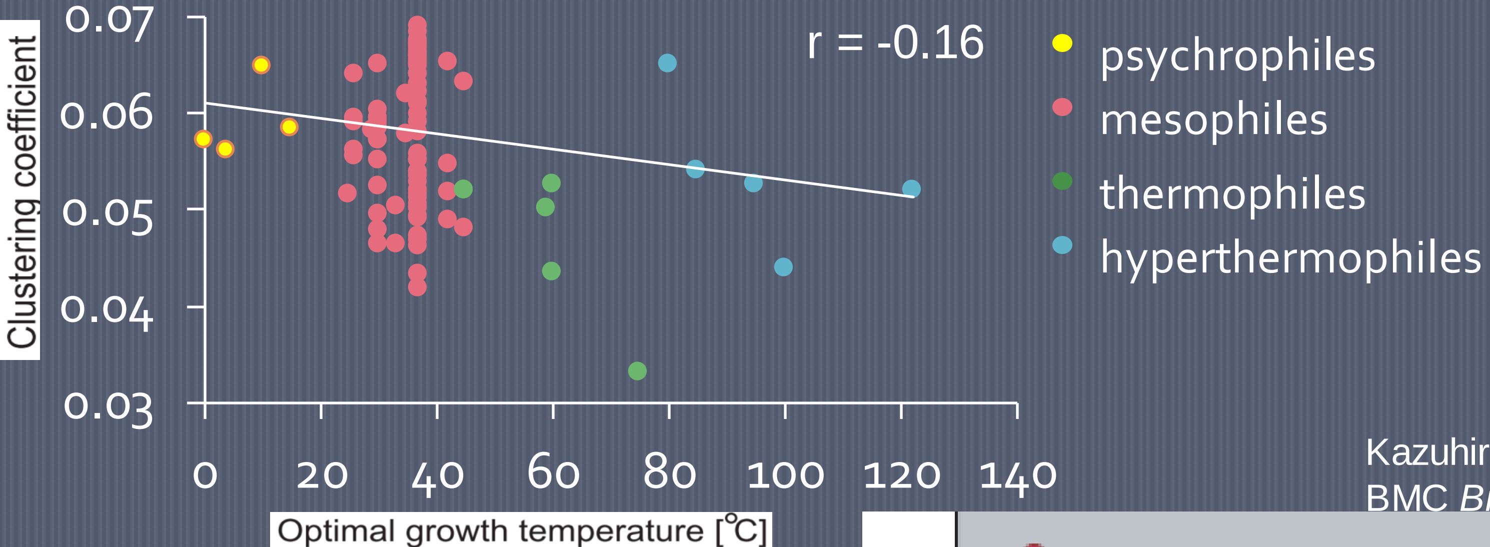
Average Clustering Coefficient Vs Temperature



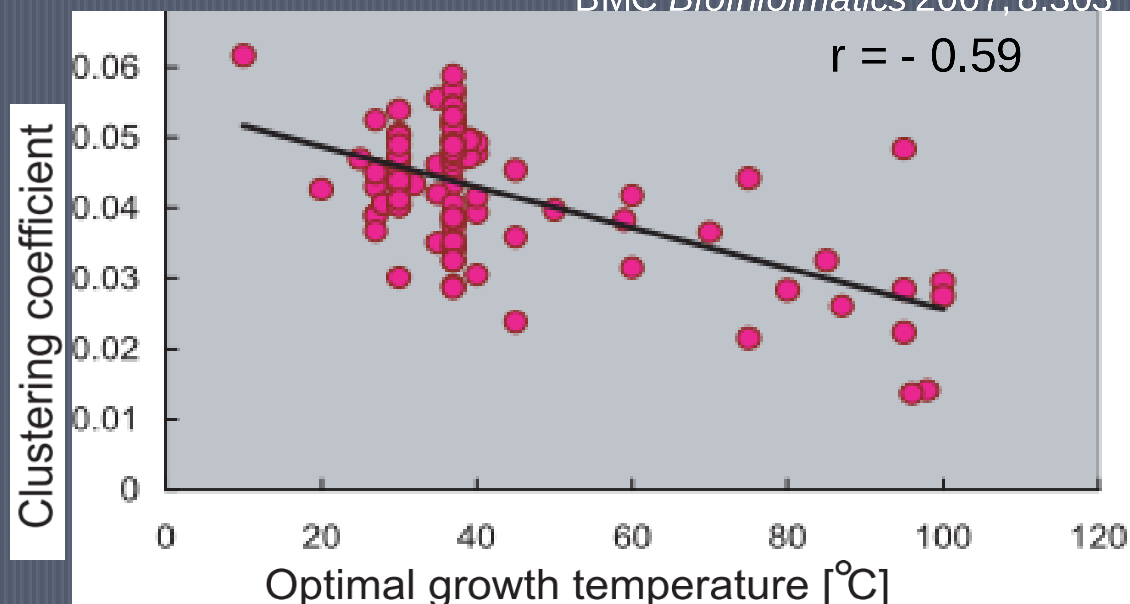
1. **There are fewer alternative paths in thermophiles and hyperthermophiles**

Results

Average Clustering Coefficient Vs Temperature



Kazuhiro Takemoto et. al.
BMC *Bioinformatics* 2007, 8:303



Reasons

1. Difference in the organisms chosen

2. I have taken only Giant Strong Component

Results

Comparison between directed and undirected network

	Pearson's Correlation coefficient	
	DIRECTED	UNDIRECTED
Clustering coefficient	-0.17	-0.16
Diameter	-0.07	0.03
Average path length	0.47	0.29
Vertices	-0.42	-0.42
Edge density	-0.25	-0.36
degree exponent	0.14	0.27

Results

Correlation in samples of comparable size

	Pearson's correlation coefficient (r)
Clustering Coefficient	-0.29585
Diameter	-0.07244
Average Path length	0.082194
Vertices	-0.65155
edge density	-0.64251
degree exponent	0.637152

Discussion

1. With present data , we can infer that density and size of gsc decrease while Degree exponent and Path length increase with temperature
2. High temp organisms have narrow range of optimal growth temperatures so they contain fewer metabolites, this might explain the reason for reduced size of g.s.c and density
3. Thermophiles and Hyperthermophiles are ancient and other groups are evolved by addition of new vertices by Gene duplication.
4. In addition, preferential attachment of new vertices to high degree nodes might explain why mesophiles have reduced path length and degree exponent

Future Work

1. c(k) – k, distribution of clustering coefficient, as a test of hierarchical structure
2. We need to look at other parameters such as assortativity, k-core analysis

Thank You

Extra

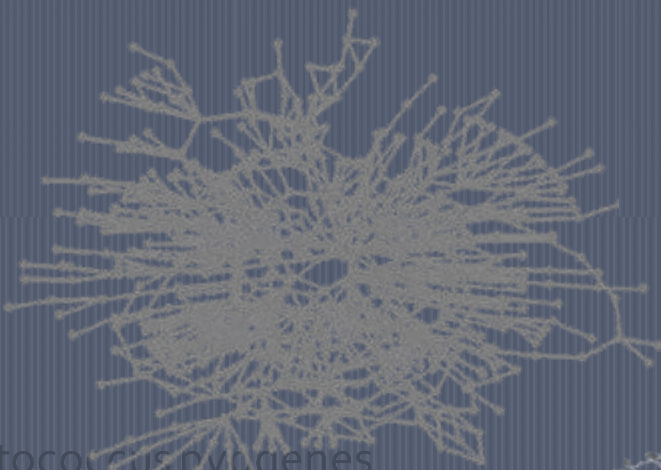
- Giant strong component also shows power law degree distribution
- Parasites have short avg path length – Ma and Zeng 2002
- Metabolic networks evolve by gene duplication – Papp et. al. Nature 2004, Diaz-Mejia et. al. Genome Biology 2007
- Preferential attachment in metabolic networks – Light et. al. BMC Bioinformatics 2005

Extra

Conclusion

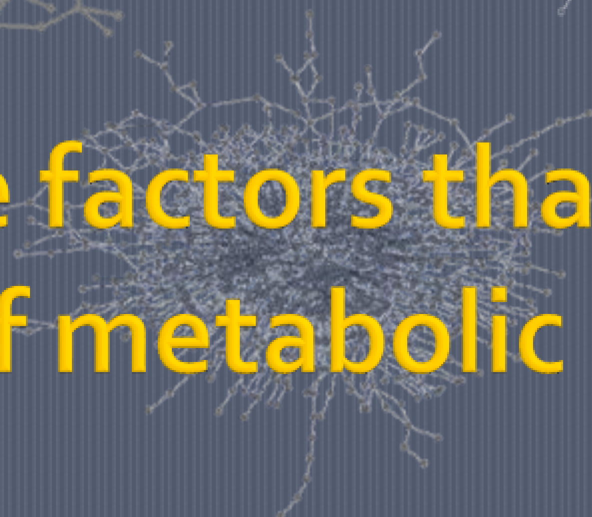
- Temperature plays important role in shaping the structure, in turn function of metabolic network

What are factors that shaped structure of metabolic networks ?

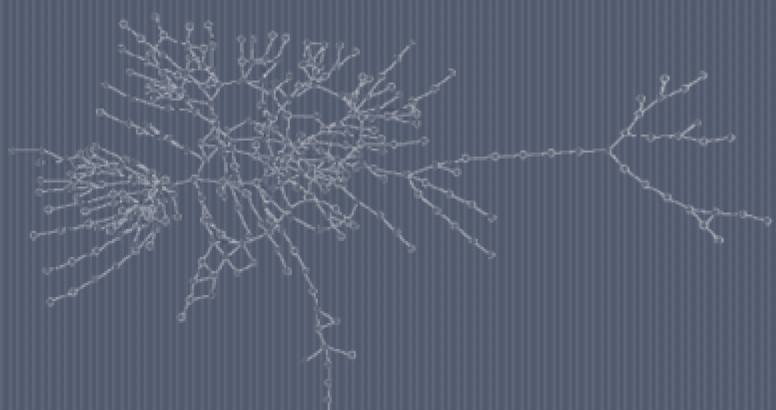


Streptococcus pyogenes

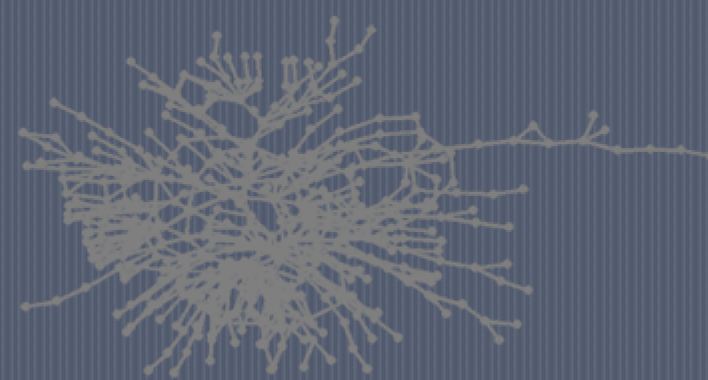
Mycoplasma genitalium



E.coli

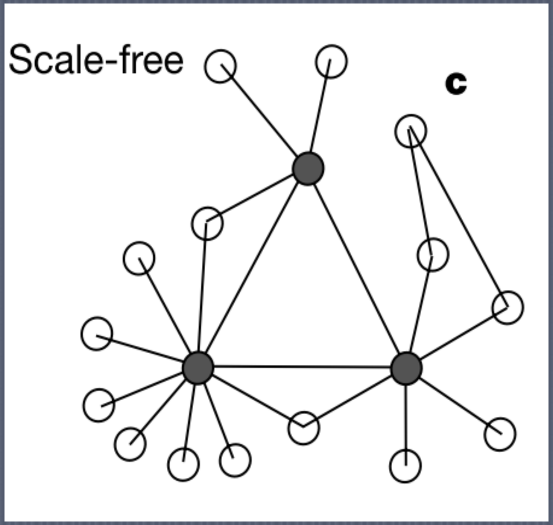


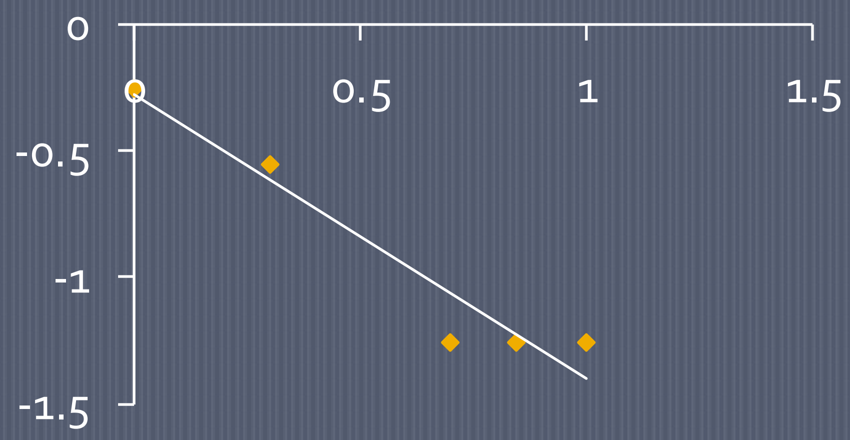
Methanopyrus kandleri



Sulfolobus acidocaldarius

- When do you call a network Scale Free ? $P(k) \sim k^{-\gamma}$





- Wiki – If a network follows power law degree distribution then it is called scale free
- If you take a fraction of the network (nodes) and plot the same graph (here degree distribution) then the slope (or gamma) remains the same