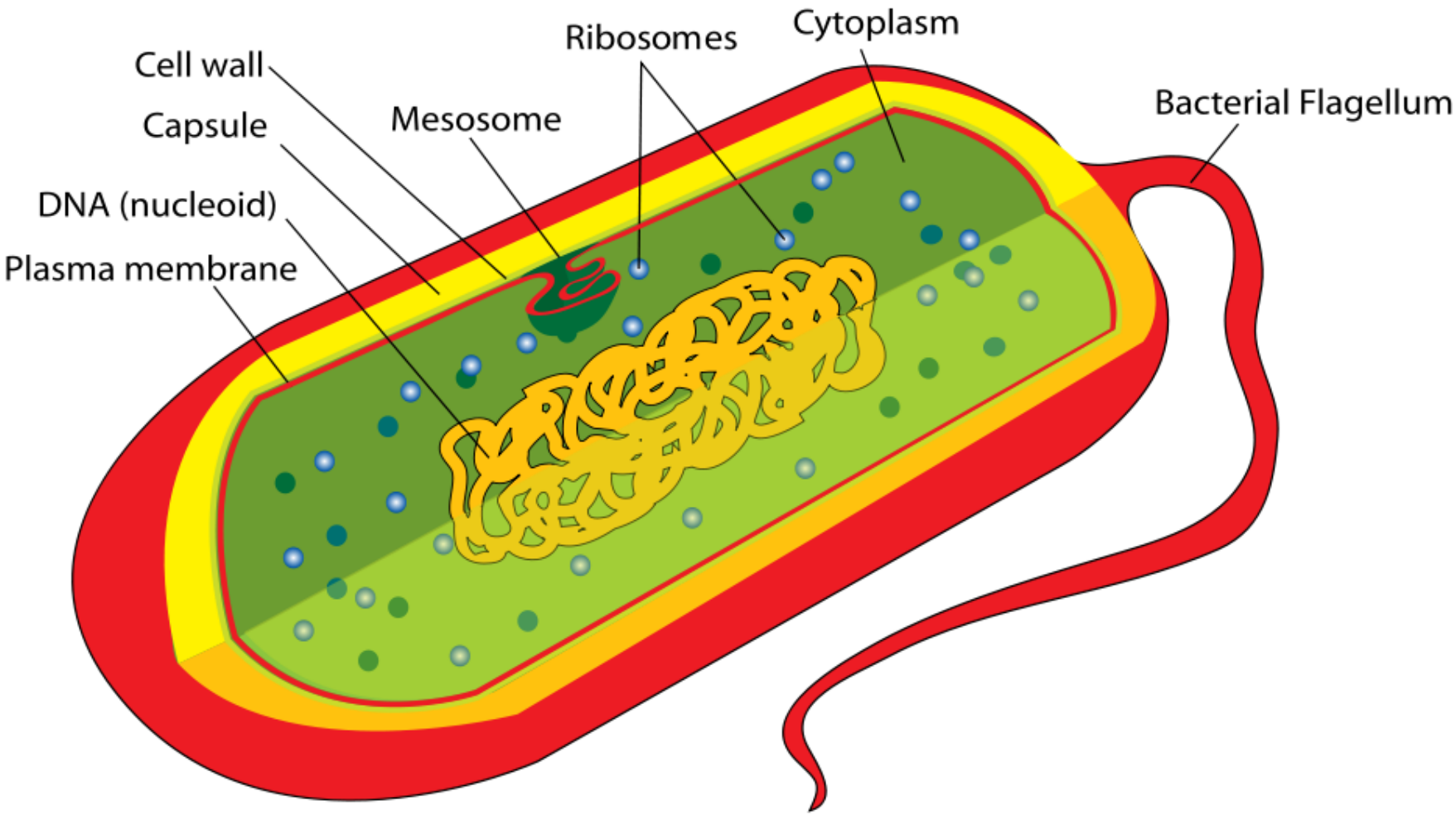
First Year Ph.D. Presentation
Constructing Metabolic Networks from Mass Spectrometry Analysis

Gordon Govan
G.M.Govan@hw.ac.uk

MACS

March 31, 2011

The Cell

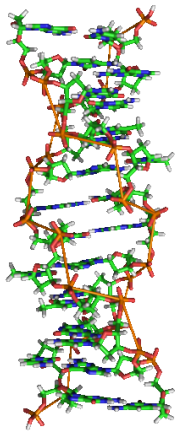


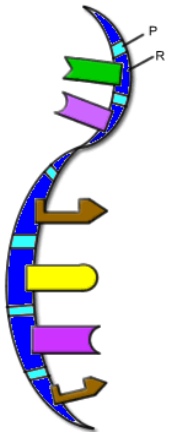
1

[1] http://commons.wikimedia.org/wiki/File:Prokaryote_cell_diagram.svg

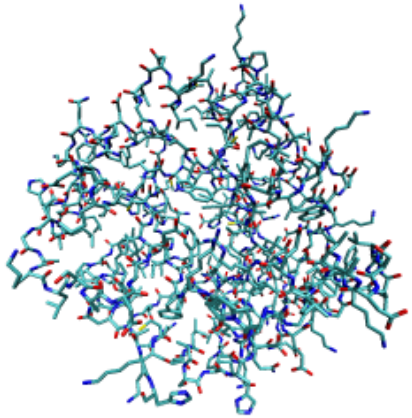
The Central Dogma of Molecular Biology

$$\text{DNA} \xrightarrow{TRANSCRIPTION} \text{RNA} \xrightarrow{TRANSLATION} \text{Proteins}$$


2



3

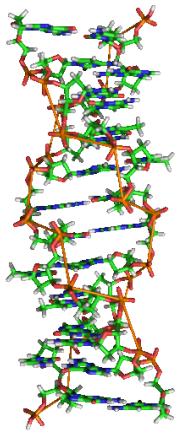

4

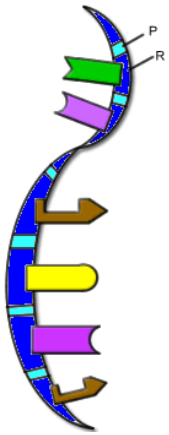
- There are approx 3 billion base pairs of DNA in the human genome.
- A Gene is the information on the DNA used to describe a protein.

[2] http://en.wikipedia.org/wiki/File:A-DNA,_B-DNA_and_Z-DNA.png
[3] http://www.biologycorner.com/bio1/DNA.html
[4] http://en.wikipedia.org/wiki/File:Proteinviews-1tim.png

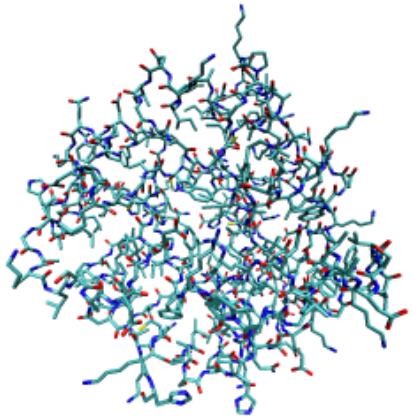
The Central Dogma of Molecular Biology

$$\text{DNA} \xrightarrow{TRANSCRIPTION} \text{RNA} \xrightarrow{TRANSLATION} \text{Proteins}$$


2



3


4

- There are approx 3 billion base pairs of DNA in the human genome.
- A Gene is the information on the DNA used to describe a protein.
- *Uses the Genetic Code*

[2] http://en.wikipedia.org/wiki/File:A-DNA,_B-DNA_and_Z-DNA.png
[3] http://www.biologycorner.com/bio1/DNA.html
[4] http://en.wikipedia.org/wiki/File:Proteinviews-1tim.png

The Genetic Code

- All the genetic information of an organism is stored in the DNA
- There are 4 bases used in DNA: G, C, A, and T
- These are *nucleic acids*
- Genes are **transcribed** from DNA to RNA, which is similar but uses the bases G, C, A, and **U**
- Information then **translated** from RNA to Proteins
- *Amino acids* are the building blocks of proteins

Metabolic Networks

- A metabolic network describes the different metabolic pathways that exist in a cell
- A metabolic pathway describes a series of linked chemical reactions that can happen within the cell.
- Chemical reactions involve
  - Substrates - the inputs
  - Enzyme - a catalyst, can be though of as a tool
  - Products - the outputs

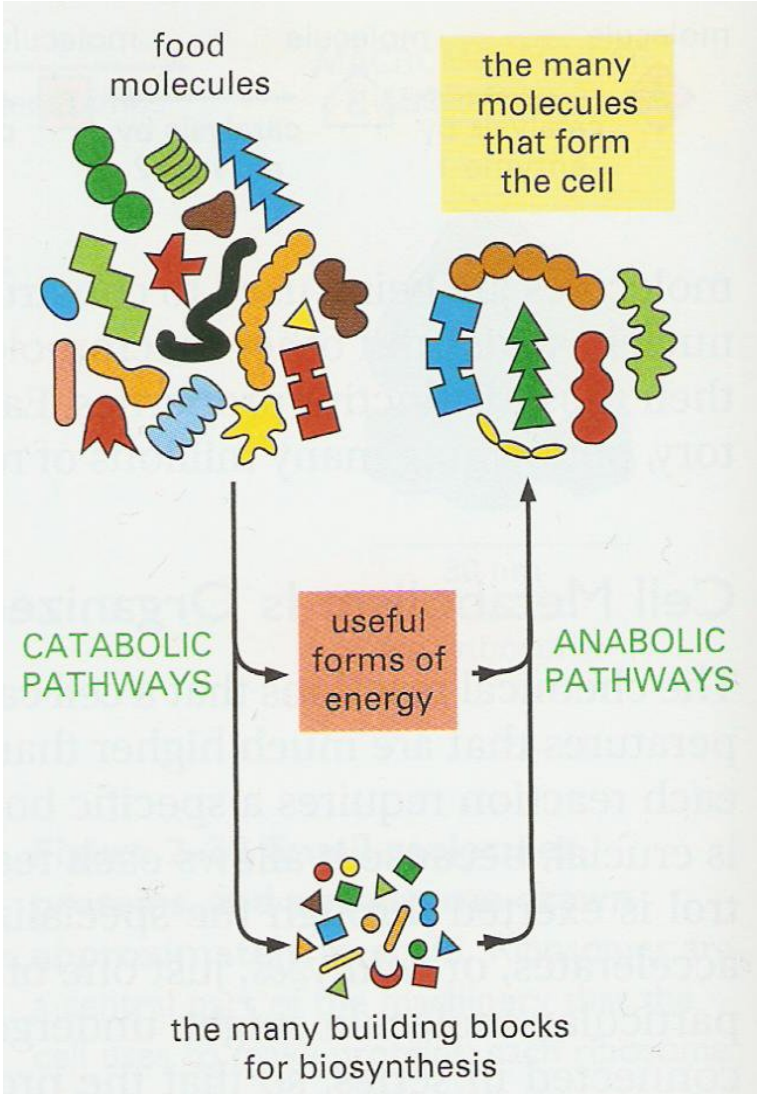
Metabolic Networks

- A metabolic network describes the different metabolic pathways that exist in a cell
- A metabolic pathway describes a series of linked chemical reactions that can happen within the cell.
- Chemical reactions involve
  - Substrates - the inputs
  - Enzyme - a catalyst, can be though of as a tool
  - Products - the outputs

$$Substrates + Enzyme \longrightarrow Products + Enzyme$$

Types of Reaction

- Two types of chemical reactions in the cell:
  - I Catabolic - release energy from food stuffs and produce small molecules
  - II Anabolic - use energy from catabolic reactions to synthesise larger molecules

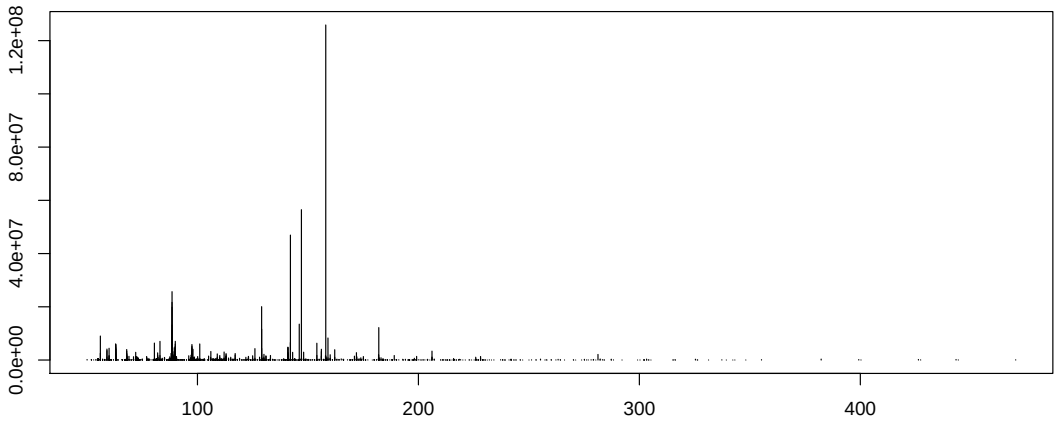


[a]

[a] Molecular Biology of the Cell, 4th edition Alberts et al.

Mass Spectrometry

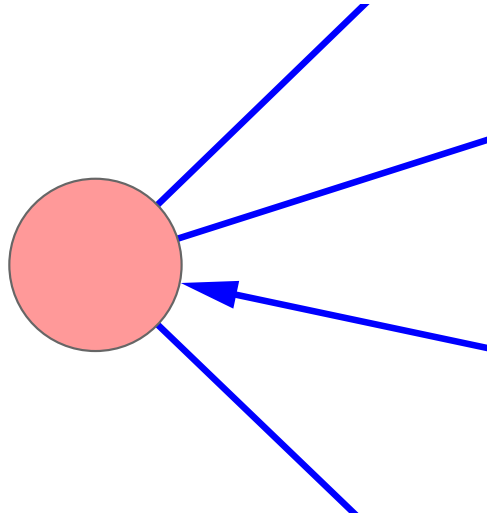
- We were given data from Rainer Brietling and Andris Jankevics in the College of Medical, Veterinary and Life Sciences at Glasgow University.
- A metabolic network constructed from the data of mass spectrometry analysis.
- Nodes were created from the peaks on this data.
- Edges were created between two nodes whose masses differed by the mass of a known metabolic transform.



Mass Spectrometry Network

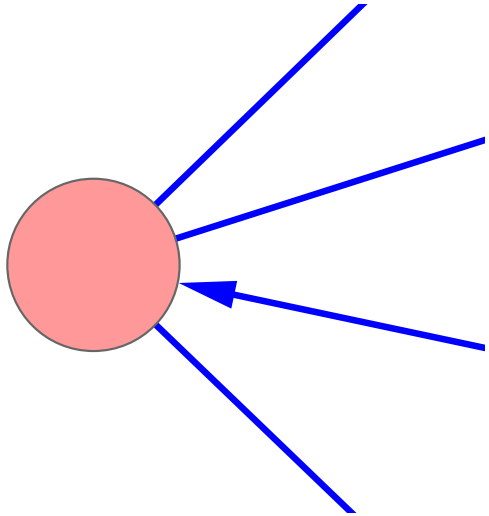
Degree Distribution

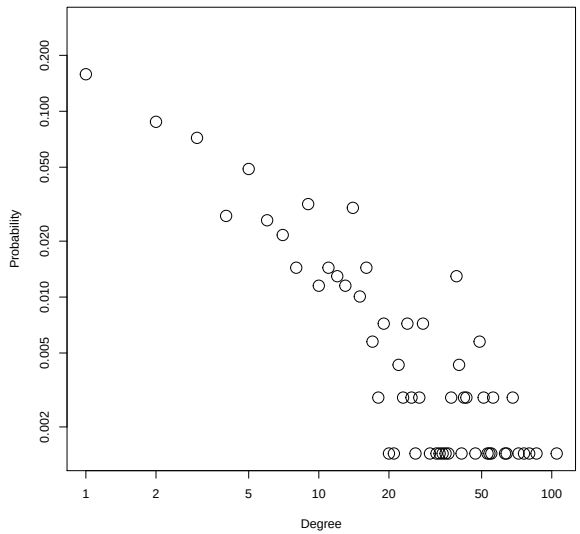
- The *Degree* of a node is the number of edges it has
- Can be counted for both incoming and outgoing edges
- Tells us how connected a node is.



Degree Distribution

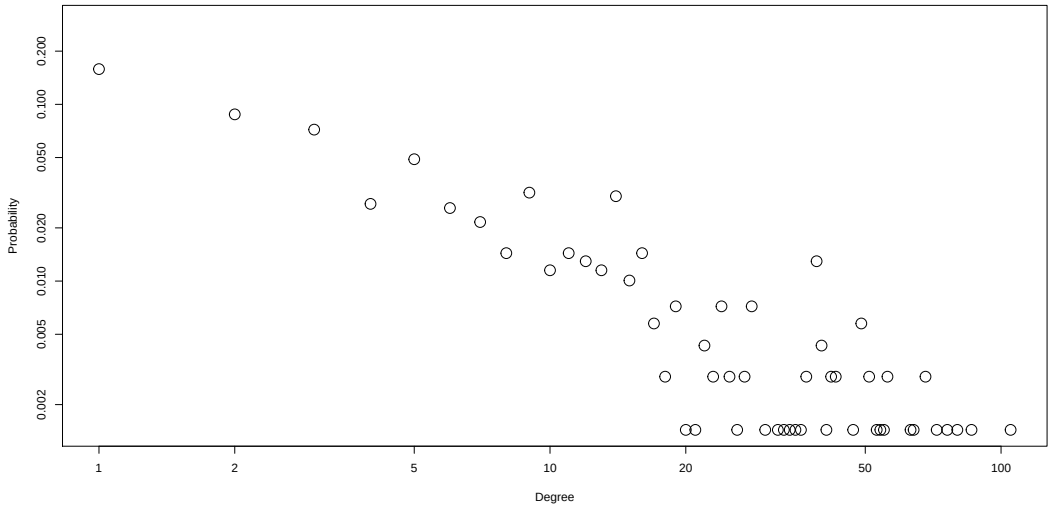
- The *Degree* of a node is the number of edges it has
- Can be counted for both incoming and outgoing edges
- Tells us how connected a node is.

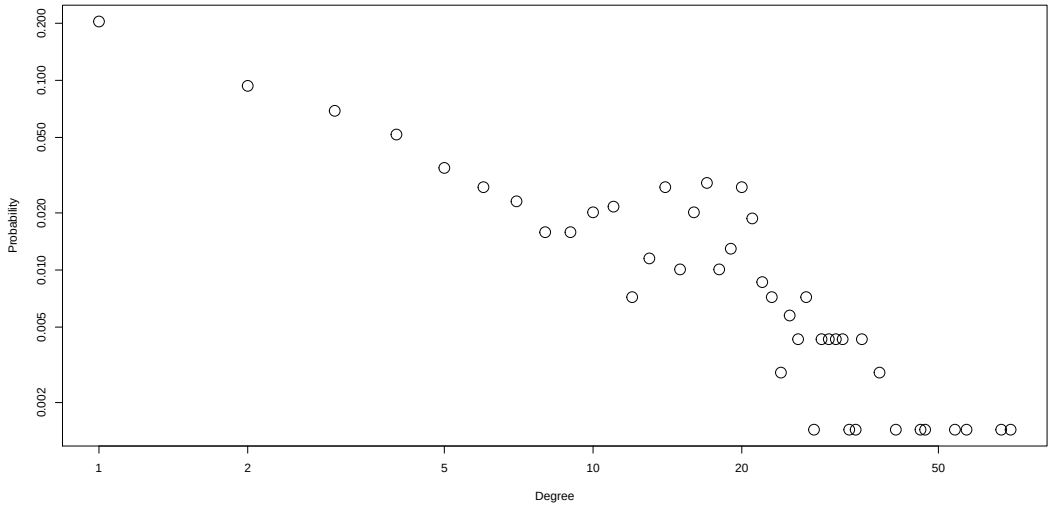




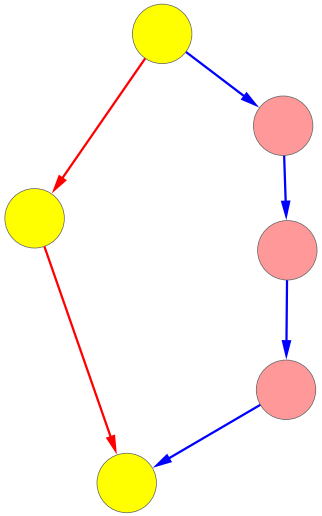
- Can be plotted as a *Degree Distribution*
- Count the degree of each node
- Count how many times each degree appears
- Divide by the number of nodes to get a probability

Degree Distribution



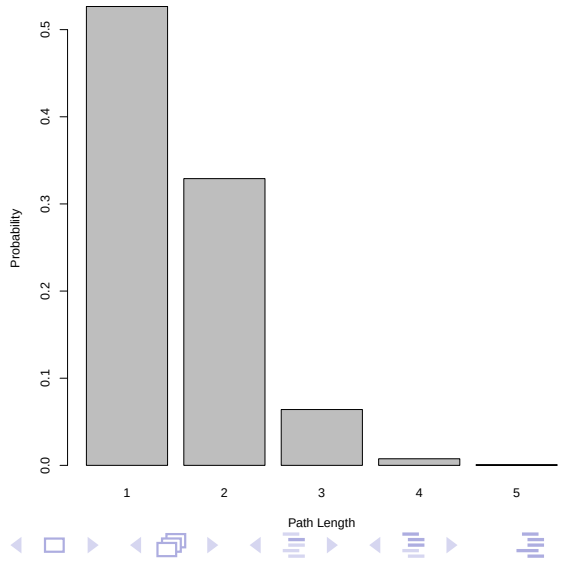


Path Lengths

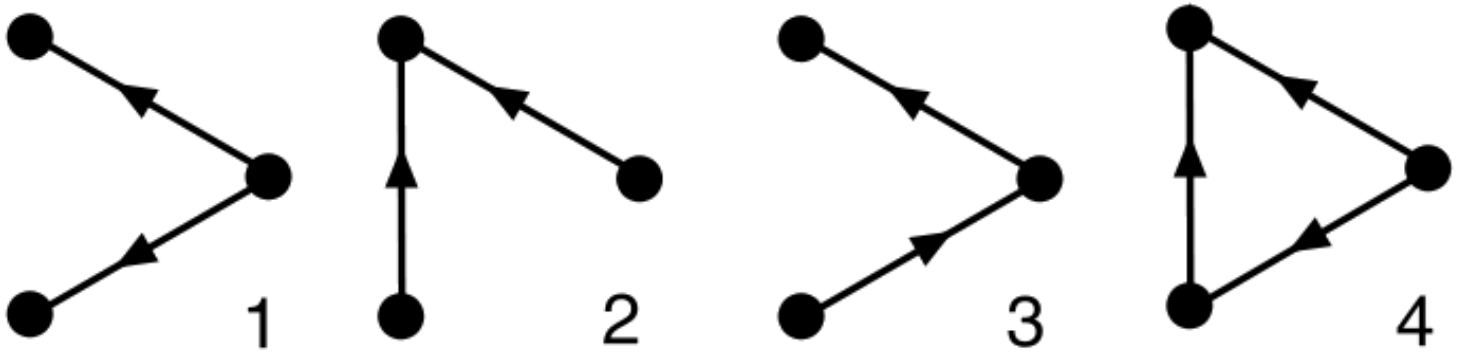


- If it is possible to reach a node from another node then we find the path length between them.
- If there is more than one path between two nodes then only count the shortest one

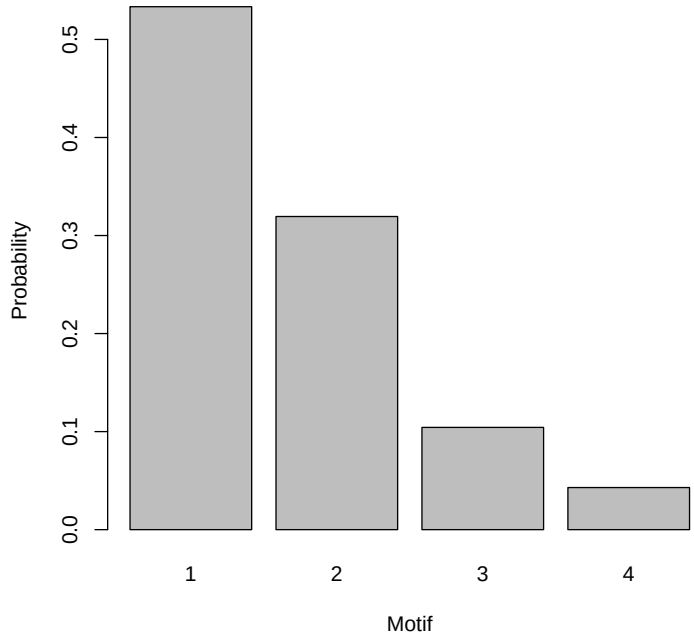
- Again we can plot a distribution
- We divide by the number of paths we counted
- This network has a longest path of 5 edges.



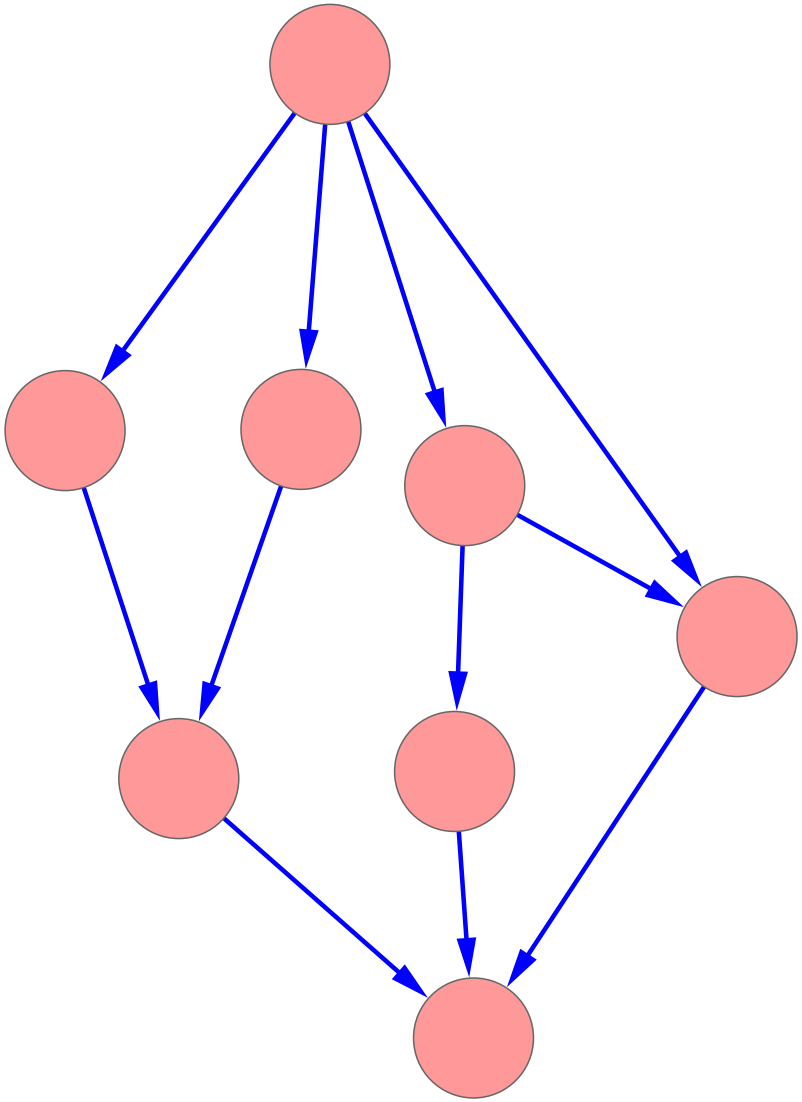
3-Node Motifs



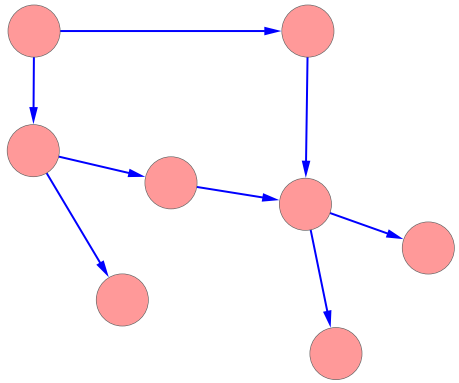
- A motif is a sub-network that is repeatedly found within larger networks
- Give an idea of the basic structural elements of a network
- 13 three-node motifs
- But only 4 in acyclic networks

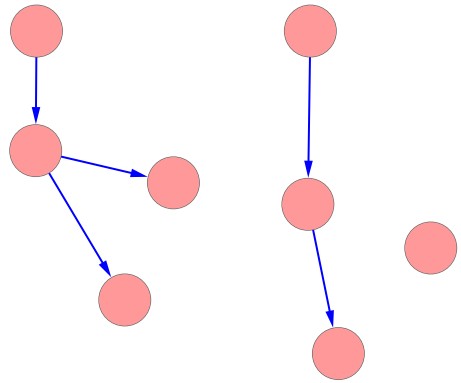


3-Node Motifs

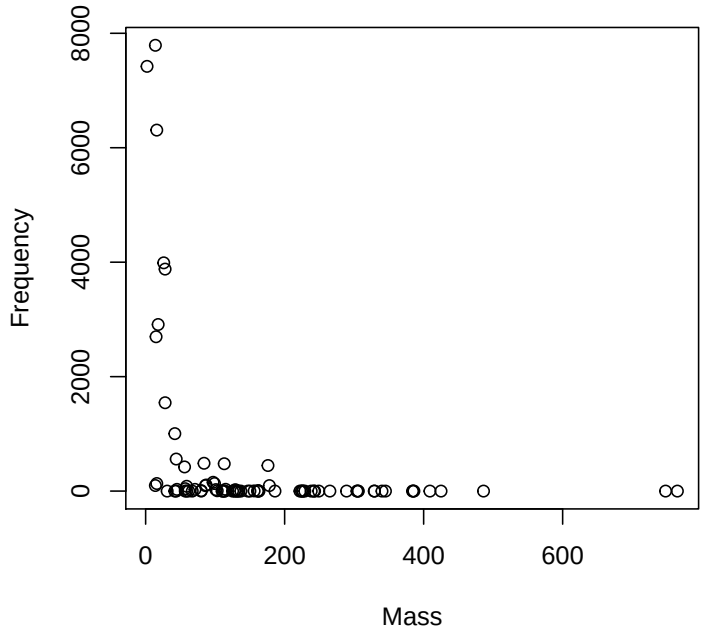


Edges



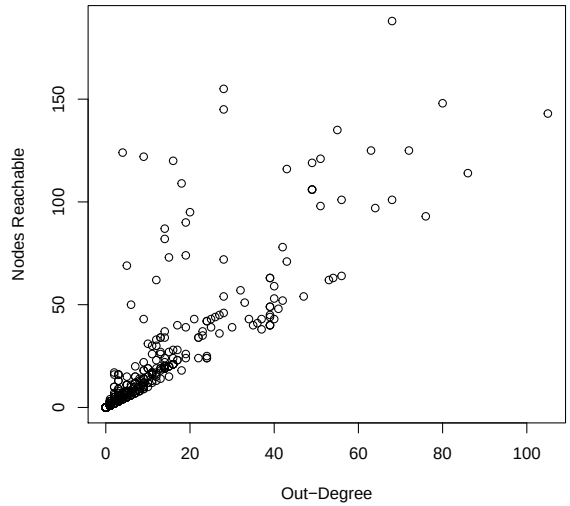


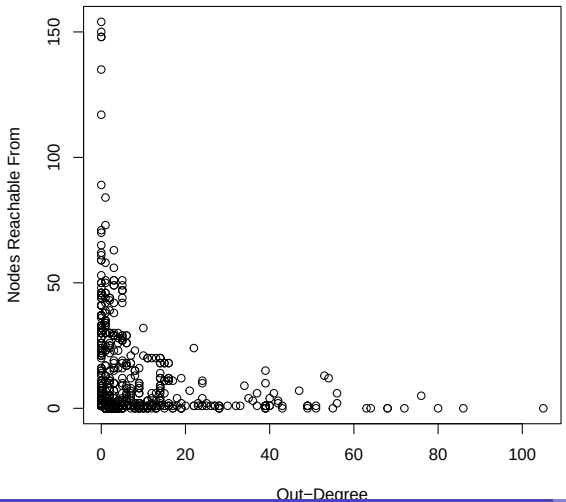
- Metabolic transforms are represented by edges
- The same transform may appear as many edges in the network
- if one of these transforms were to be removed then the topology of the network could change considerably
- The frequency with which these transforms occur is important



Nodes Reachable

- For each node we can count the number of nodes that it can be reached from it
- And the number of nodes that it is reachable from
- These are plotted with the out-degree on the x-axis





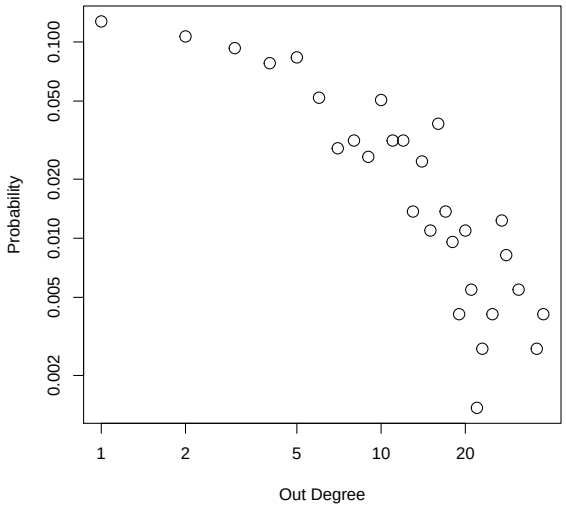
- Shows that the out-degree doesn't tell us about how many things it can reach
- But it does tell us that there is a limit on the number of nodes that it is reachable from

Simulated Mass Spectrometry Network

- The masses were taken from a database of metabolic networks
- These masses were treated as the output from the mass spectrometry
- The masses became nodes, and edges were made between them if they differed by the mass of a known metabolic transform.

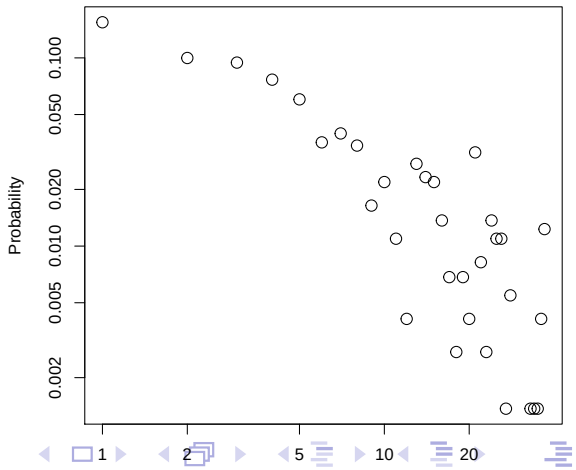
Simulated MS Network

SMS - Degree Distributions

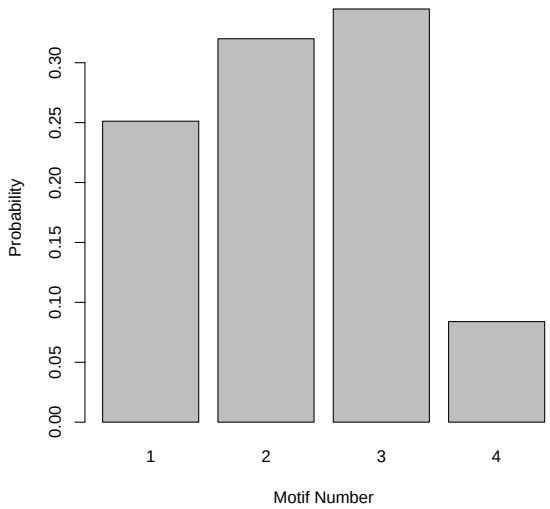


- The degree distribution is similar to the mass spectrometry network
- But there is a narrower range of degrees

- Yet the average degree (both in and out) of the two networks are very close together
- Neither of the degree distributions follow a power law for this network

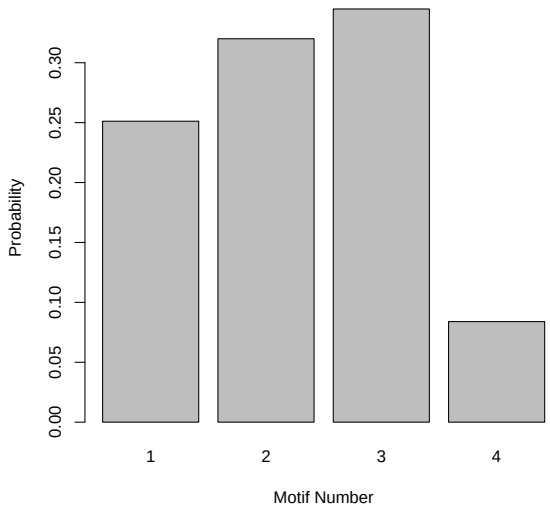


SMS - Motifs and Path Length



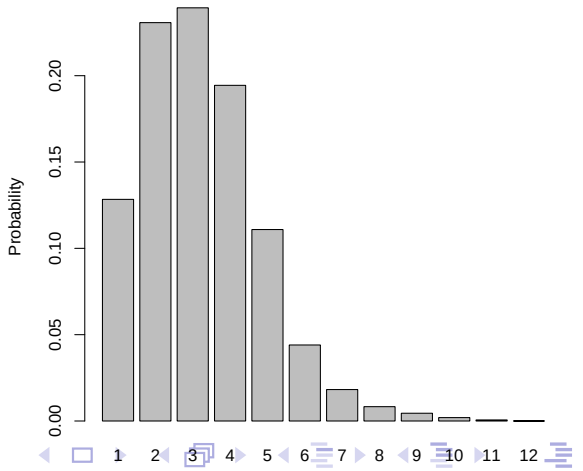
- Motifs 3 and 4 are have a higher likelihood of appearing in the simulated network
- May be because of the shorter range of the degrees
- Or the longer path lengths

SMS - Motifs and Path Length

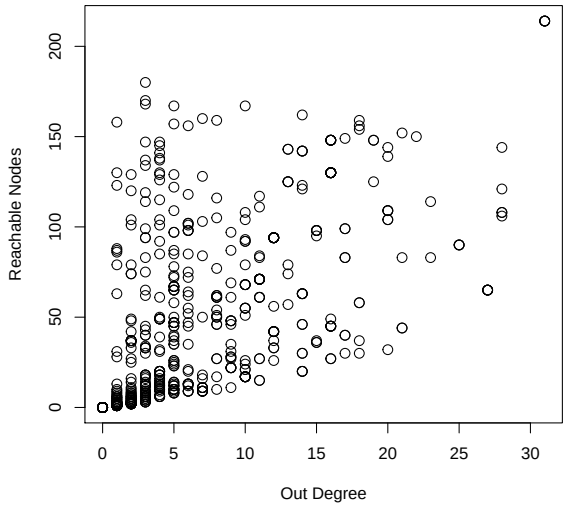


- Motifs 3 and 4 are have a higher likelihood of appearing in the simulated network
- May be because of the shorter range of the degrees
- Or the longer path lengths

- There is a great difference in the path lengths
- The longest path here is 12
- The average path length is 3 times that of the mass spectrometry network.

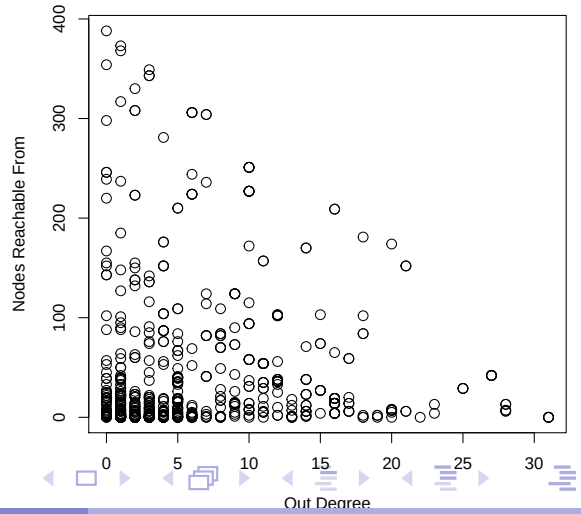


SMS - Reachable Nodes



- Again the out-degree tells us nothing about the number of nodes reachable

- But here it also tells us much less about the number of nodes that a node is reachable from
- Yet the distributions for reachable and reachable from look very similar to the mass spectrometry network



Results

- Networks are very different
- The differences between the path length and motif distributions most pronounced
- Thought to have been caused by the *noise* of the mass spectrometry

Results

- Networks are very different
- The differences between the path length and motif distributions most pronounced
- Thought to have been caused by the *noise* of the mass spectrometry

What could have been done next?

- Networks from different stages of the the life-cycle of a cell
- Networks from different cell types in a multi-cellular organism.
- Networks from different organisms

Results

- Networks are very different
- The differences between the path length and motif distributions most pronounced
- Thought to have been caused by the *noise* of the mass spectrometry

What could have been done next?

- Networks from different stages of the the life-cycle of a cell
- Networks from different cell types in a multi-cellular organism.
- Networks from different organisms
  - From each of the three kingdoms
  - Eukaryote
  - Prokaryote
  - Archea

Thanks for listening.

Any questions or complaints?