Finding Biologically Plausible Complex Network Topologies with a New Evolutionary Approach for Network Generation

Gordon Govan, Jakub Chlanda, David Corne, Alex Xenos, Pierluigi Frisco

School of Mathematical and Computing Sciences
Heriot-Watt University, Edinburgh, UK

12-07-13 – EVOLVE 2013 - Leiden University

Summary: What it is all about

- Increasing use of networks
- Networks models used to model networks
- Complex models require more parameters
- We use an EA to find parameters for a network model
- Compares favourably against trial-and-error approach

Motivation:

- Networks are used to model many biological processes
  - Metabolic pathways
  - Neural networks
  - Protein-protein interactions
  - Gene regulation

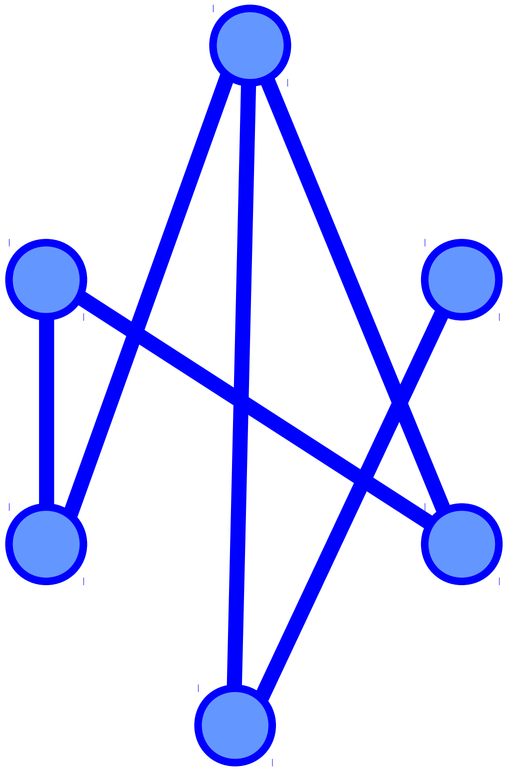
- Network models used to bring insight to development and behaviour of empirical networks

- Network models limited in the features that they can capture

Network Models: **Erdős-Rényi**

- Randomly create edges between nodes

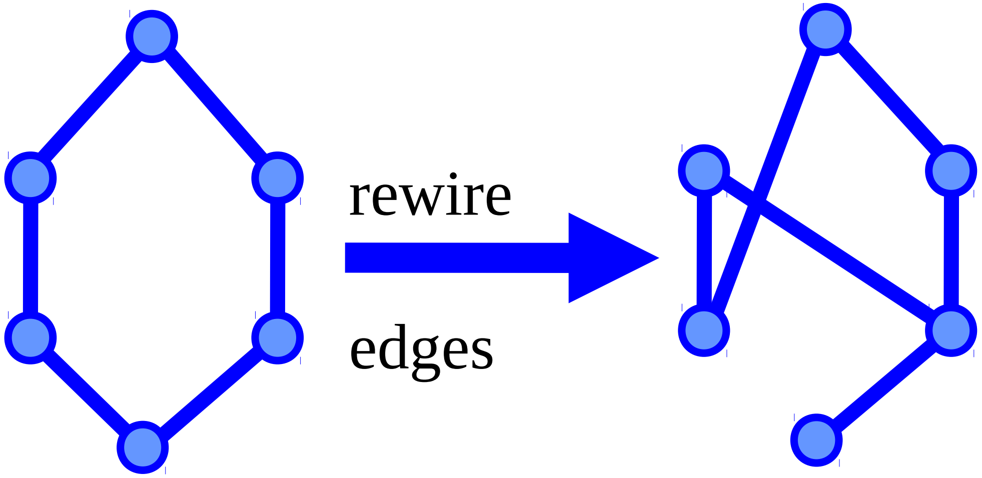
- Generated networks have a low average path length and low clustering coefficient



Network Models: **Watts-Strogatz**

- Start with a regular network, then randomly rewire edges

- Generated networks have a low average path length and high clustering coefficient

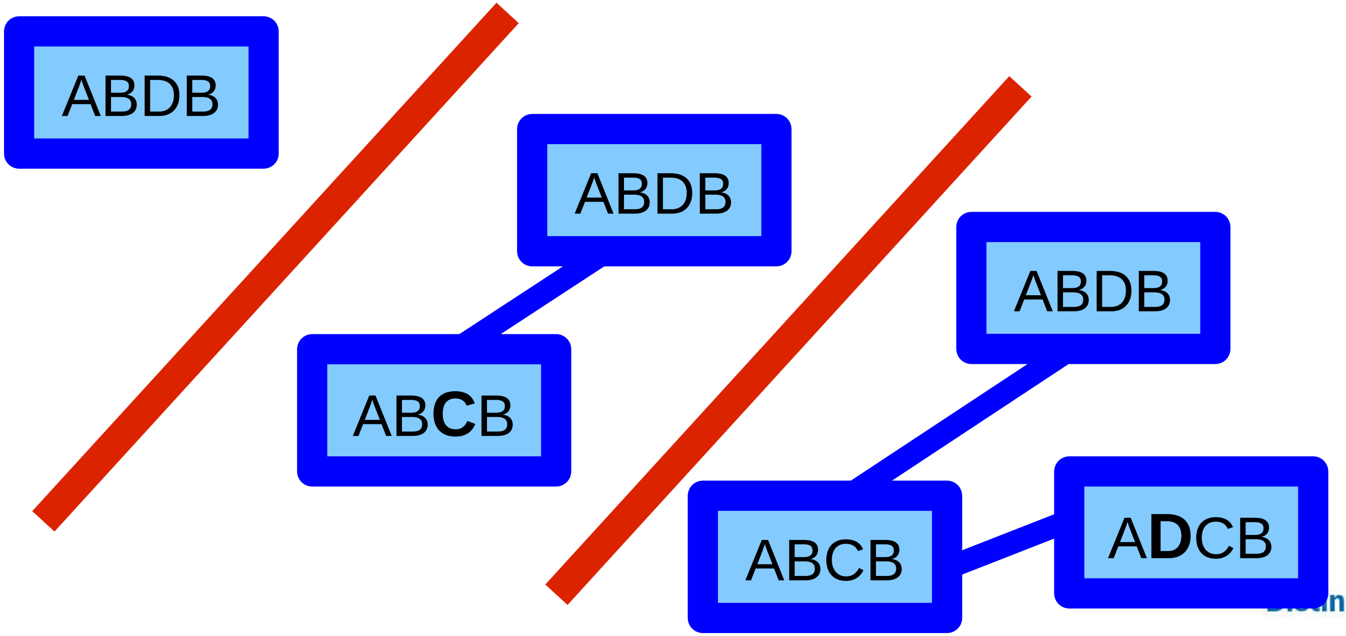


Ring Lattice

Watts-Strogatz

Network Models: Structured Nodes

- Nodes have a structure
- Creation of new nodes and edges based on node structure



ABDB

ABDB

AB**C**B

ABDB

ABCB

A**D**CB

Network Models: Structured Nodes

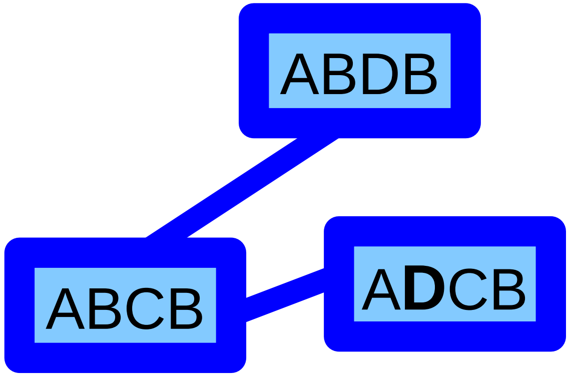
- Pick an existing node

ABCB

- Mutate its structure to create a new node

A**D**CB

- Add node to network based on a distance measure



- Repeat until you have the desired number of nodes

Generating Networks:

- How do we generate networks from network models
- We target specific features
- Select parameters which generate networks with those features
- Method used changes with complexity of model

Generating Networks: ER model

- ER model has two parameters
  - Number of nodes
  - Probability of an edge between them

- Target the number of nodes and the average degree

- p_{edge} = avg. degree / (no. nodes – 1)

- Easy to calculate

Generating Networks: WS model

- WS model has three parameters
  - Number of nodes
  - Number of neighbours connected to
  - Probability to rewire an edge

- Target
  - Number of nodes
  - Average degree
  - Average path length
  - Average clustering coefficient

- No simple formula to use

Generating Networks: WS model

- First two parameters – number of nodes & average degree
- Third parameter – more difficult
- Use hill climbing to get a good balance

Generating Networks: SN model

- Many parameters

- Number of nodes the only obvious one

- Other parameters can either be about
  - Creating nodes
  - Connecting nodes together

- Previously done through trial-and-error

- Is there a better way

Generating Networks: **SN model**

- Use an Evolutionary Algorithm (EA)
- Evolve a set of input parameters
- Run them through the SN model to generate networks
- Take measurements from these networks and compare them to a set of target values

Generating Networks: SN model

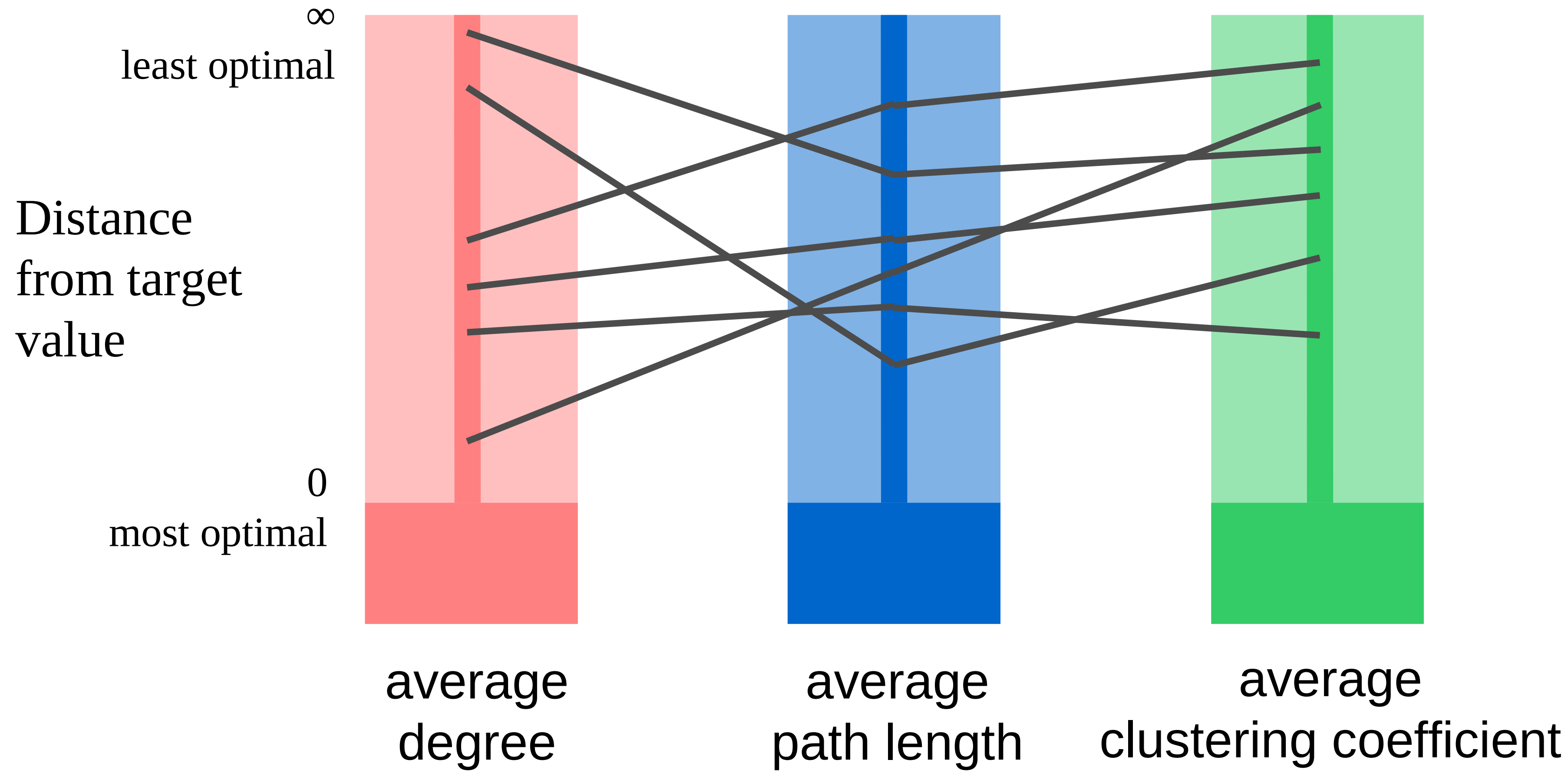
- Target values for average degree, path length and clustering coefficient

- But changing one of the parameters changes more than one target value

- Makes this a multi-objective optimisation problem

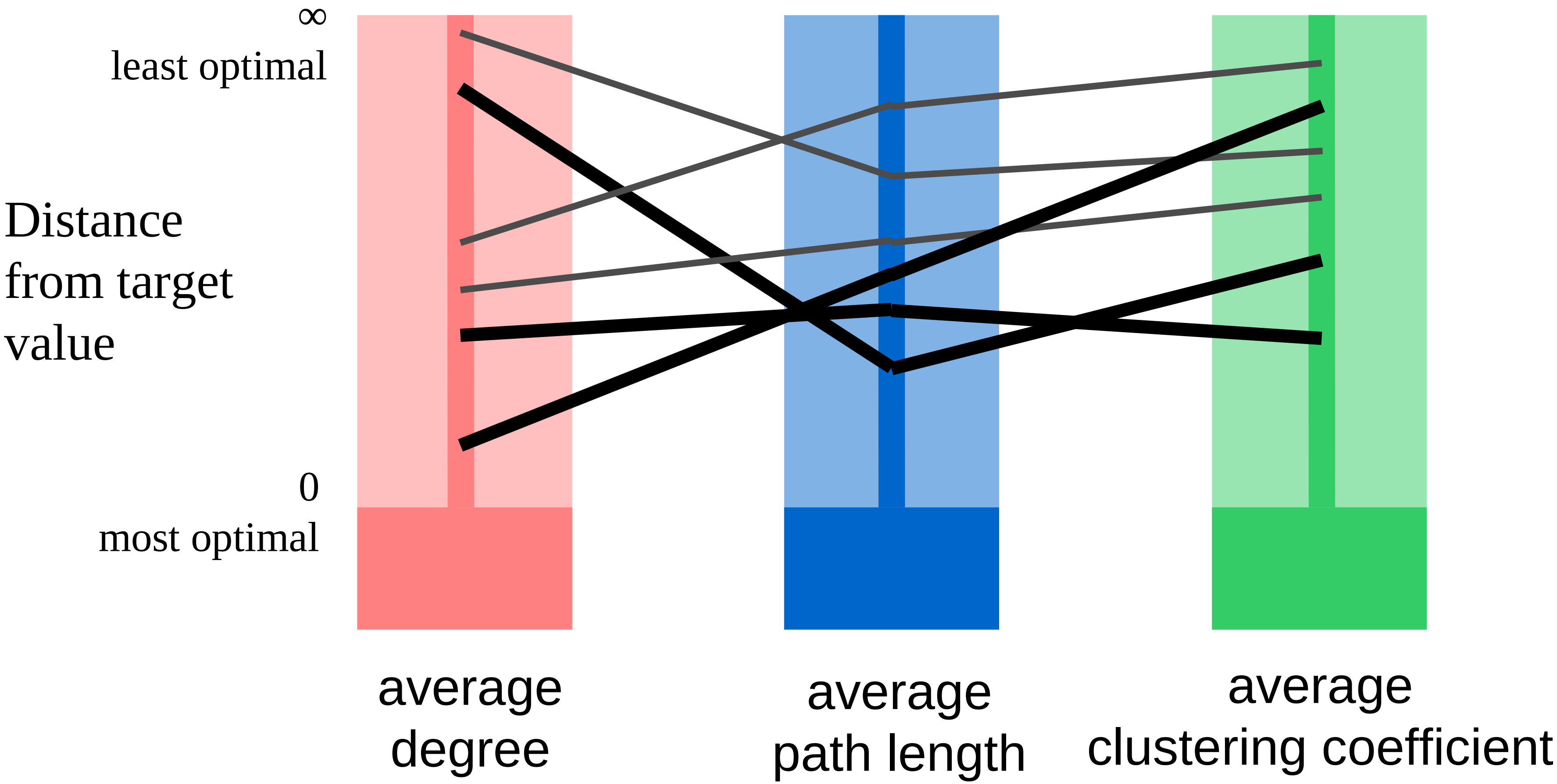
Generating Networks: SN model

- Want to get the generated network to match as closely as possible the empirical network on 3 measurements

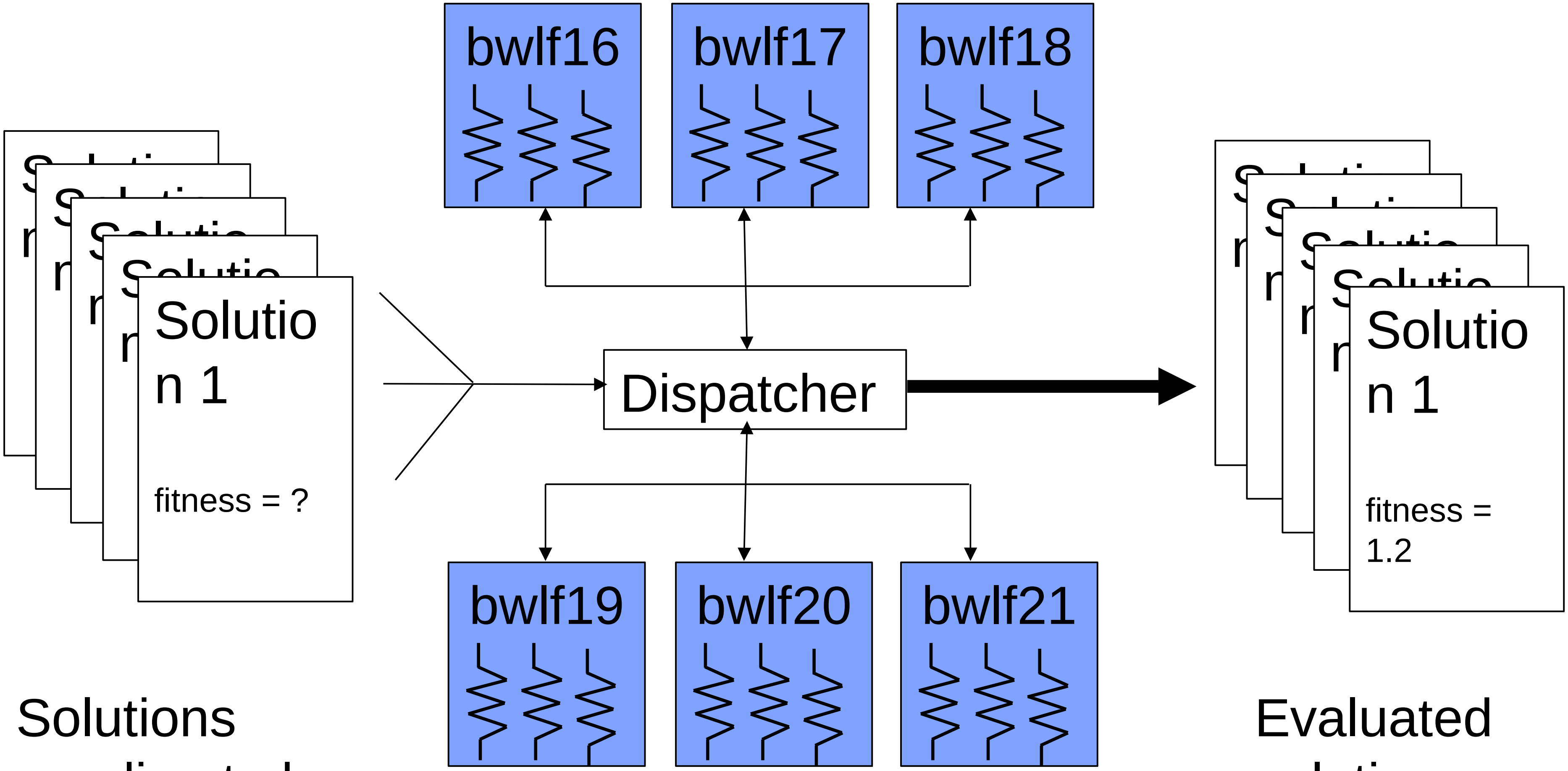


Generating Networks: **SN model**

- Want to get the generated network to match as closely as possible the empirical network on 3 measurements



Generating Networks: Parallelisation



bwlf16

bwlf17

bwlf18

Solution 1

fitness = ?

Dispatcher

Solution 1

fitness = 1.2

bwlf19

bwlf20

bwlf21

Solutions needing to be evaluated

Evaluated solutions

Results: Three empirical networks

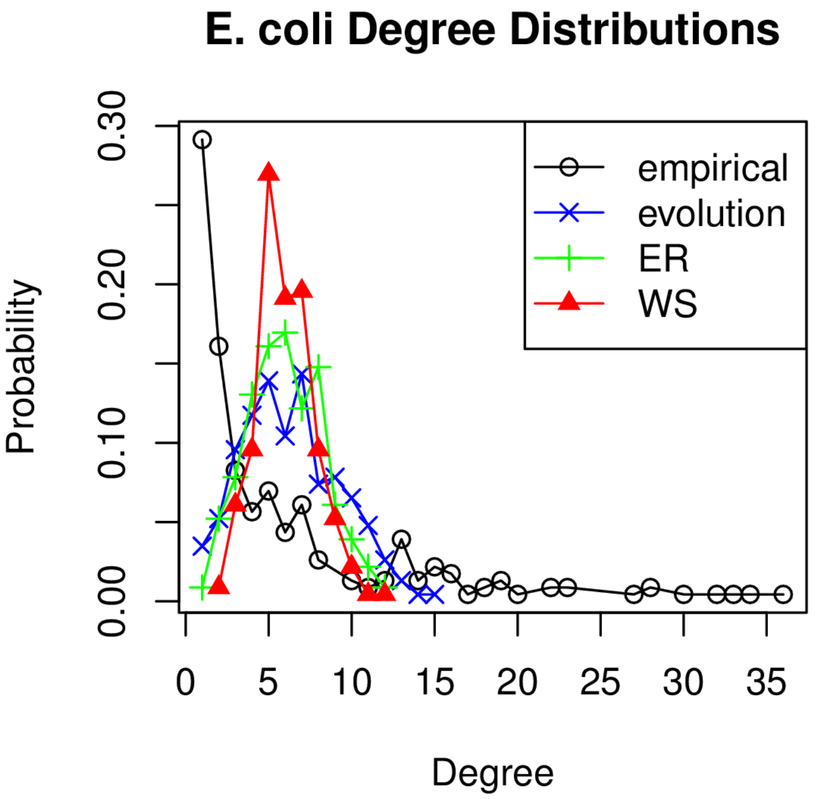
- Three empirical networks were used, with different topological features

Network	Nodes	Edges	Degree	Path length	Clustering coefficent
E. coli	230	1390	6.04	3.78	0.22
S. cerevisiae	3186	130234	40.87	3.564	0.85
MRSA	2321	37864	16.31	3.95	0.33

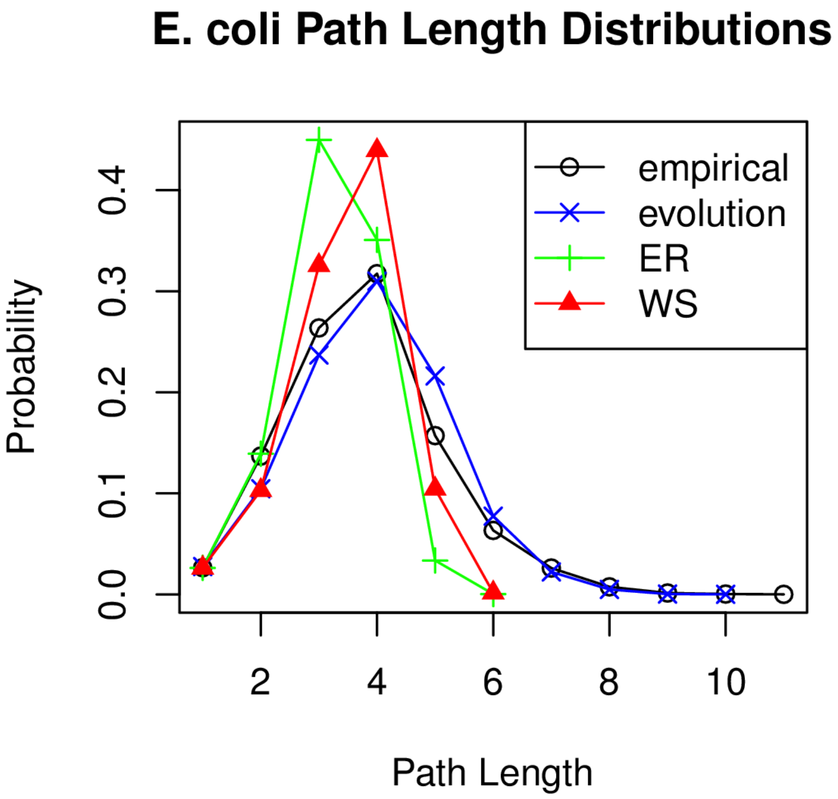
Results: *E. coli*

Network	Degree	Path length	Clustering coefficent
E. coli	**6.04**	**3.78**	**0.22**
ER	6.02	3.22	0.002
WS	6.00	3.53	0.22
SN: trial and error	6.03	3.85	0.26
SN: EA	6.60	3.77	0.20

Results: ***E. coli***



E. coli Degree Distributions

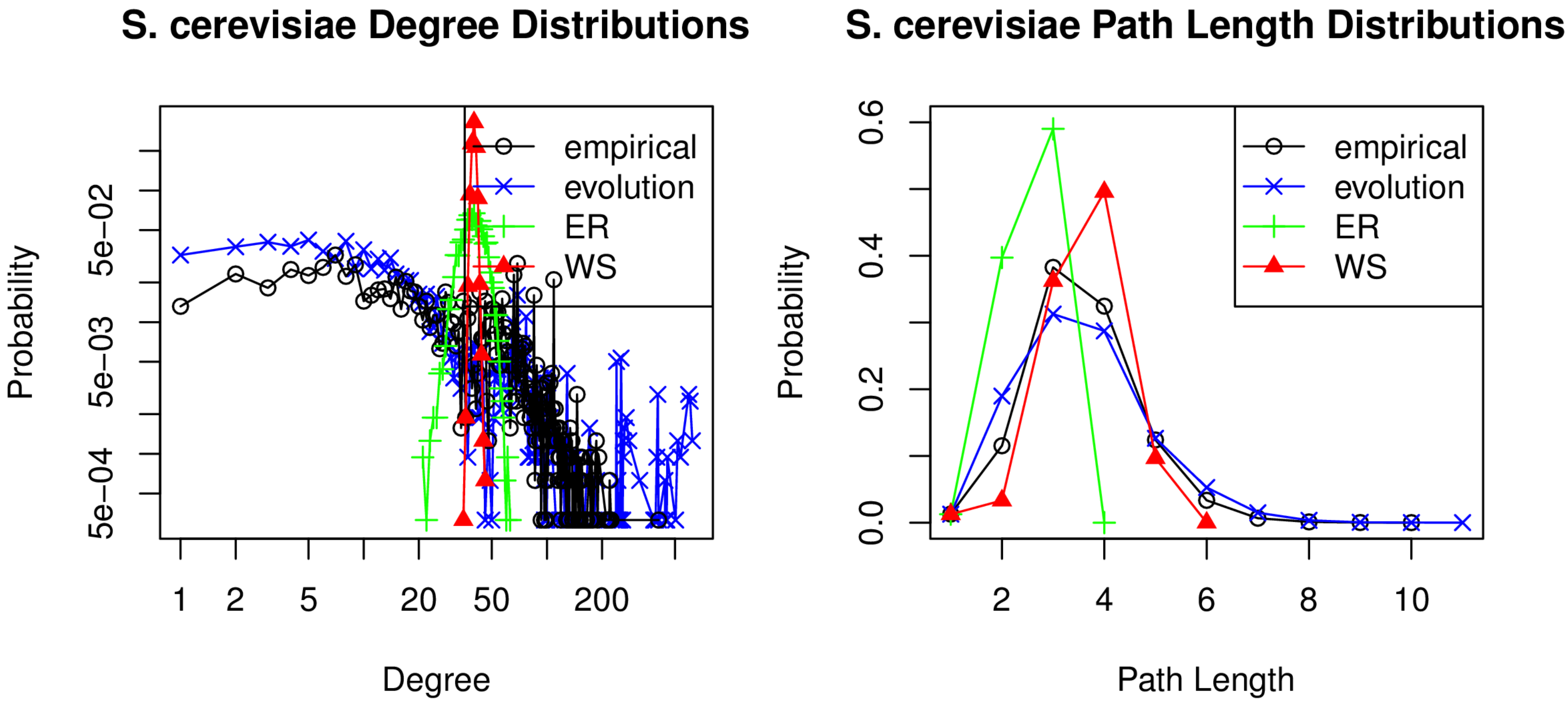


E. coli Path Length Distributions

Results: *S. cerevisiae*

Network	Degree	Path length	Clustering coefficent
S. cerevisiae	**40.87**	**3.564**	**0.85**
ER	40.70	2.57	0.013
WS	40.00	3.67	0.69
SN: trial and error	41.50	3.04	0.44
SN: EA	48.50	3.31	0.80

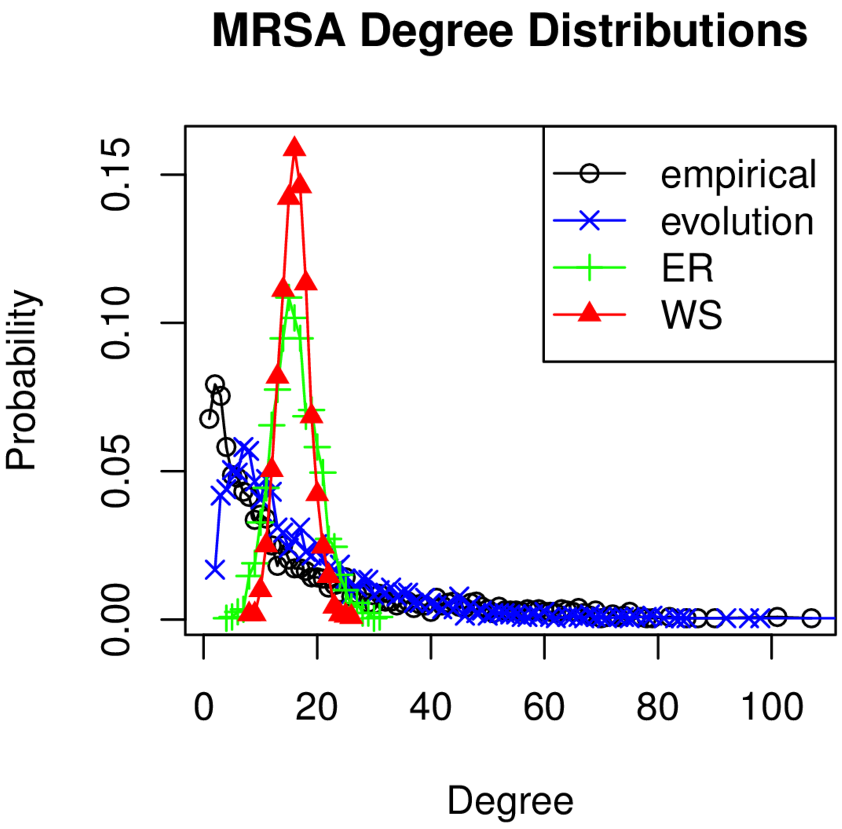
Results: *S. cerevisiae*

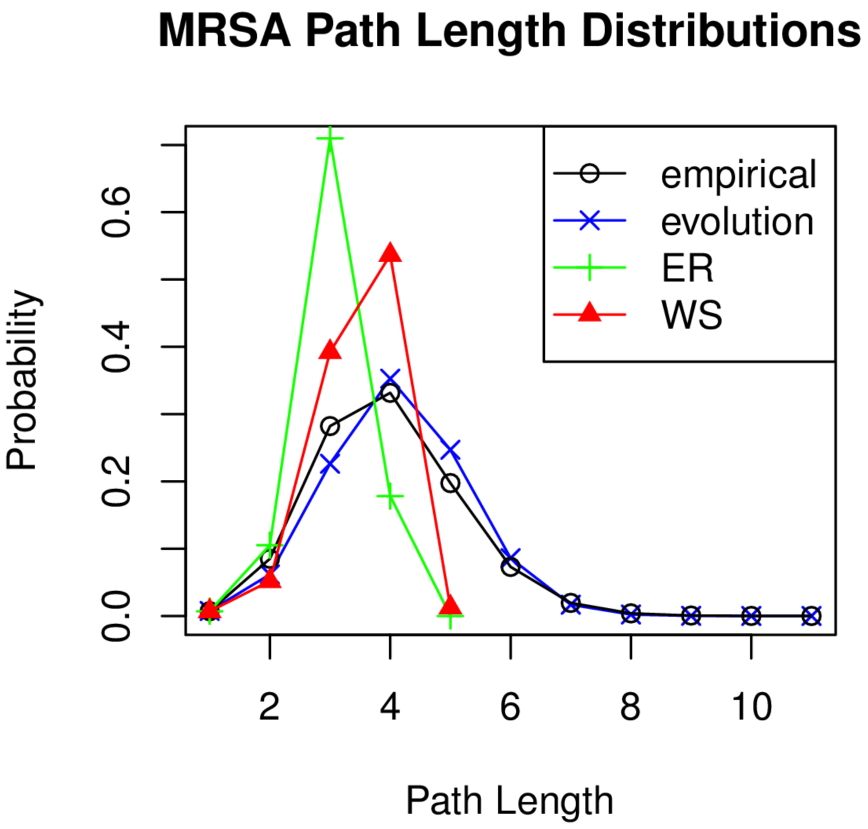


Results: MRSA

Network	Degree	Path length	Clustering coefficent
MRSA	**16.31**	**3.95**	**0.33**
ER	16.30	3.05	0.007
WS	16.00	3.51	0.33
SN: EA	18.51	4.07	0.47

Results: MRSA


MRSA Degree Distributions


MRSA Path Length Distributions

Results: Parameters

E. coli

Prob to add = 0.0
Prob to delete = 0.0
Prob to duplicate = 0.0
Prob to mutate = 1.0
Max distance = 1
Unit distance = 1
Initial node = CCBAC
Alphabet = A,B,C,D

S. cerevisiae

Prob to add = 1.0
Prob to duplicate = 0.0
Prob to delete = 0.0
Prob to mutate = 0.0
Max distance = 1
Unit distance = 5
Initial node = ABCBACDDADCCADBAA
Alphabet = A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,P

MRSA

Prob to add = 0.002
Prob to delete = 0.002
Prob to duplicate = 0.002
Prob to mutate = 0.994
Max distance = 2
Unit distance = 1
Initial node = ABCABABAABBCA
Alphabet = A,B,C

Conclusions:

- Using EA is much easier than trial-and-error
- Creates novel solutions
- Very good at matching the path length and clustering coefficient
- Further work – match other features, including distributions.

All code and data at
www.macs.hw.ac.uk/~gg32/projects/modelnetwork

Thanks for listening,
Any questions...?

Gordon Govan
gmg8@hw.ac.uk