# Mixed and time-varying models for network formation

**Naomi Arnold**, Raul Mondragón, Richard Clegg

4 July 2019

Queen Mary
University of London

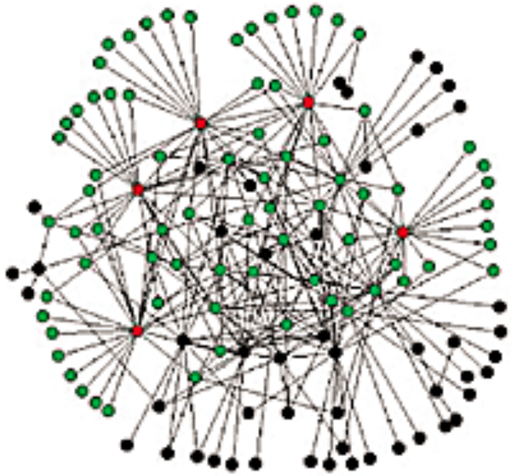
# How do networks grow?

# How do networks grow?

**Emergence of Scaling in Random Networks**

Albert-László Barabási* and Réka Albert

**33759 Citations**



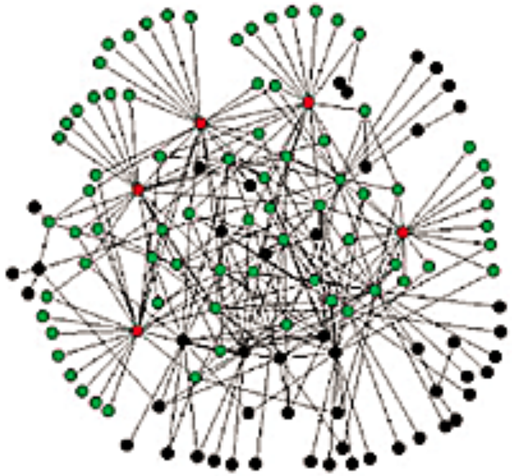
**Preferential Attachment**

# How do networks grow?

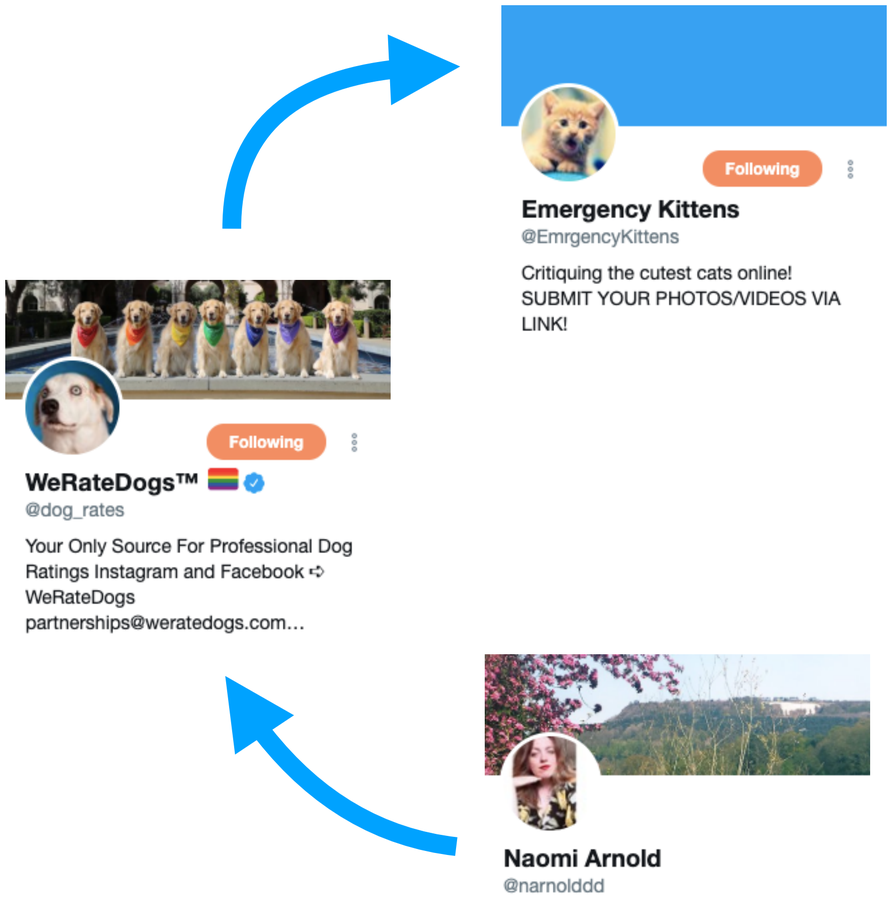
**Emergence of Scaling in Random Networks**

Albert-László Barabási* and Réka Albert

**33759 Citations**



**Preferential Attachment**



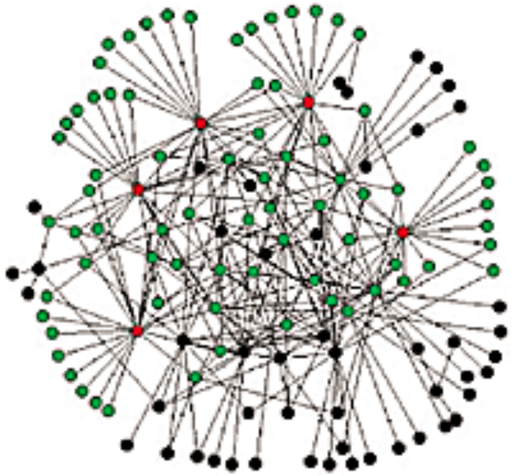

**Triangle Closure**

# How do networks grow?

**Emergence of Scaling in Random Networks**

Albert-László Barabási* and Réka Albert

**33759 Citations**



**Preferential Attachment**

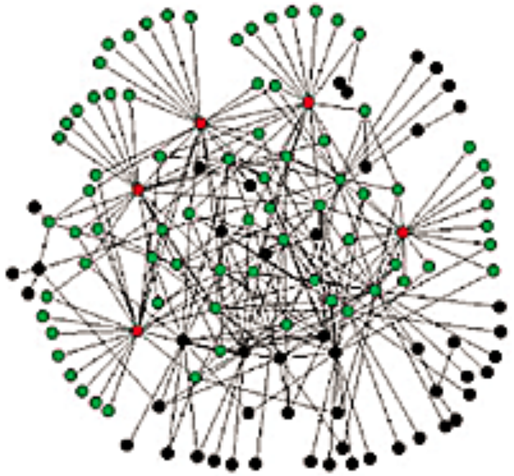

**Triangle Closure**

# How do networks grow?

**Emergence of Scaling in Random Networks**

Albert-László Barabási* and Réka Albert

**33759 Citations**



**Preferential Attachment**


**Triangle Closure**


**Random Meeting**

# Our hypothesis

**The model best describing growth of a network comprises a mixture of mechanisms…**

**… and this mixture may change over time.**

# The Cosener's Example

**Arrival coffee: Random interaction**

**Time**

# The Cosener's Example

**Arrival coffee: Random interaction**


**Second Coffee session: Preferential Attachment/Triangle Closure**

**Time**

# The Cosener's Example

**Arrival coffee: Random interaction**

*Have you met my student?*

*Cool talk!!*

*I'm buying a round!*

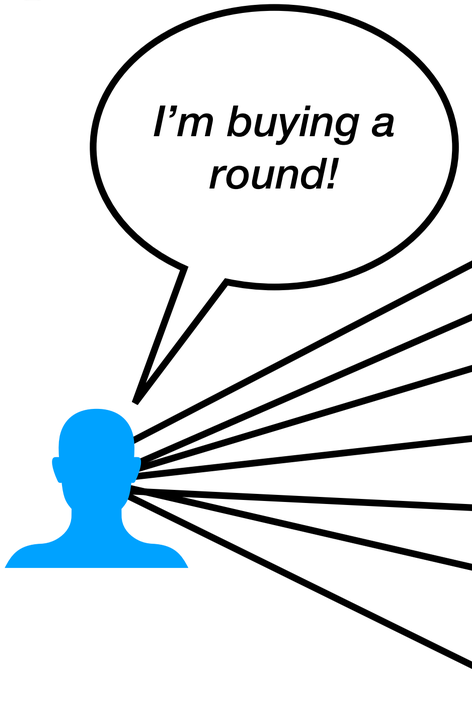
**Second Coffee session: Preferential Attachment/Triangle Closure**

**Time**

# The Cosener's Example

**Arrival coffee: Random interaction**


**Second Coffee session: Preferential Attachment/Triangle Closure**




**Evening Social: Preferential Attachment?**

**Time**

# Actual example: Enron email network

# Actual example: Enron email network




**Enron scandal: multiple well-documented events in company's downfall**

# Actual example: Enron email network


**Enron scandal: multiple well-documented events in company's downfall**

```
3385 2935 1006874609
3385 2944 1006874609
3385 2951 1006874609
3385 2954 1006874609
3385 2956 1006874609
3385 2957 1006874609
3385 2975 1006874609
3385 2990 1006874609
3385 2991 1006874609
3385 2999 1006874609
3385 3024 1006874609
3385 3053 1006874609
3385 3387 1006874609
```

Sender Receiver UNIX

**Corpus of emails between employees handed over for investigation**

# Actual example: Enron email network


**Enron scandal: multiple well-documented events in company's downfall**

?

**Were events in the scandal reflected in the evolution of the email network?**

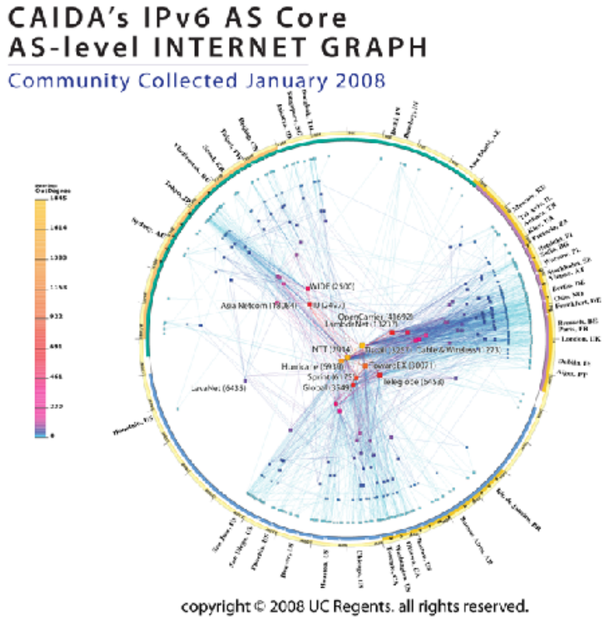
```
3385 2935 1006874609
3385 2944 1006874609
3385 2951 1006874609
3385 2954 1006874609
3385 2956 1006874609
3385 2957 1006874609
3385 2975 1006874609
3385 2990 1006874609
3385 2991 1006874609
3385 2999 1006874609
3385 3024 1006874609
3385 3053 1006874609
3385 3387 1006874609
```

Sender Receiver UNIX

**Corpus of emails between employees handed over for investigation**

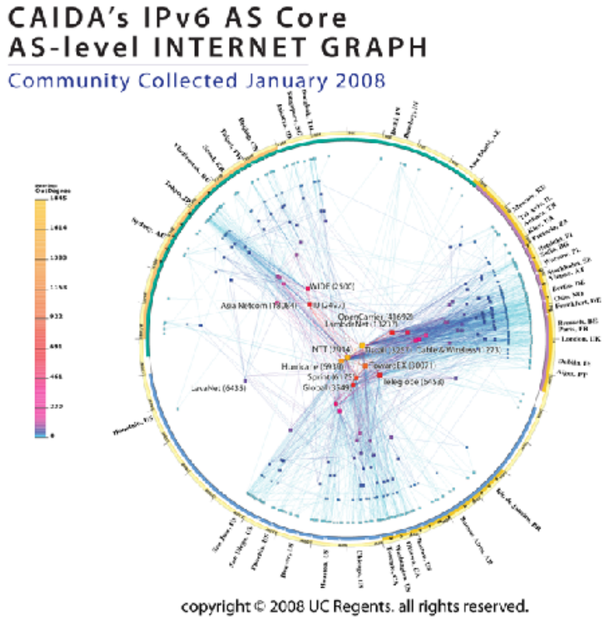
# How do we traditionally choose a model?

# How do we traditionally choose a model?

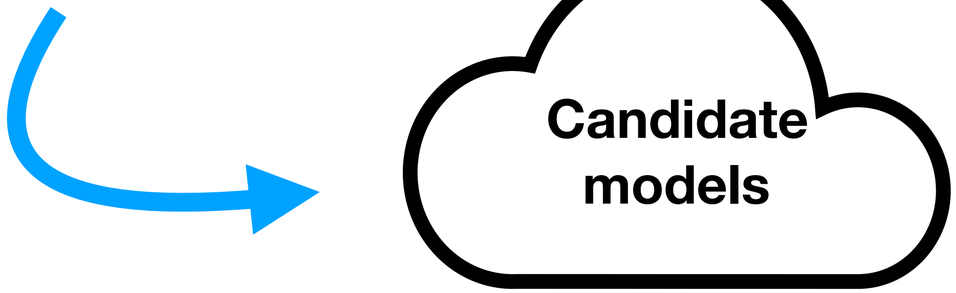



**Network we want to model**

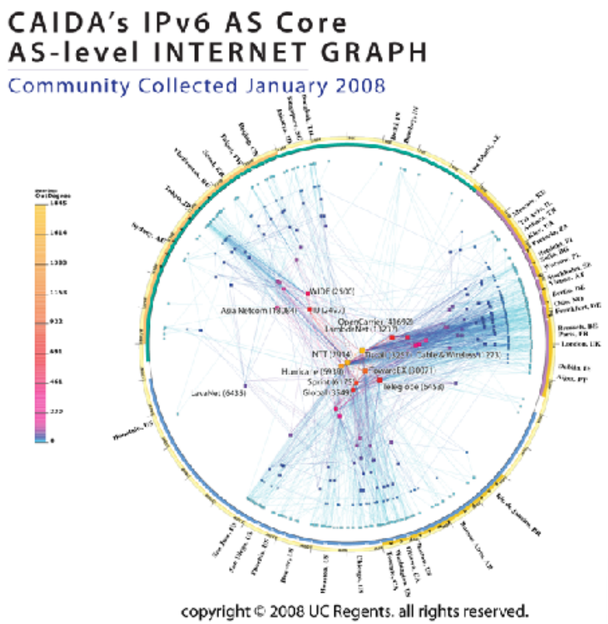
# How do we traditionally choose a model?



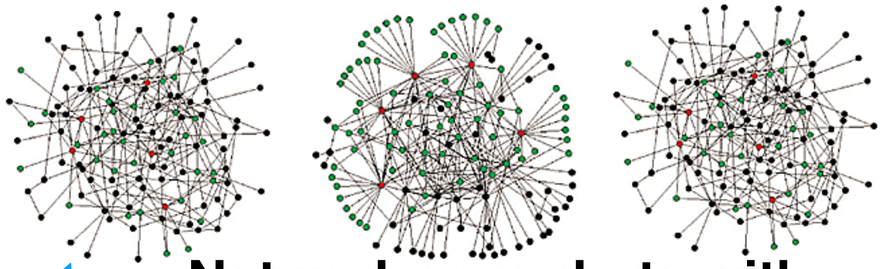

**Network we want to model**



# How do we traditionally choose a model?




Network we want to model

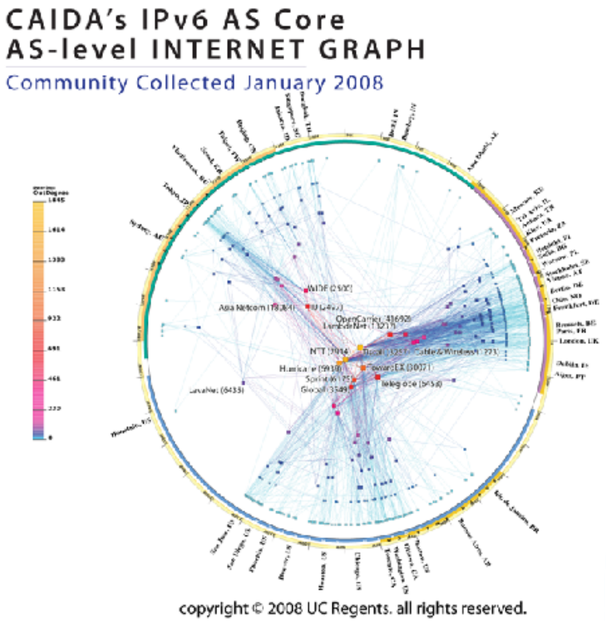


Network snapshots with same #nodes/links

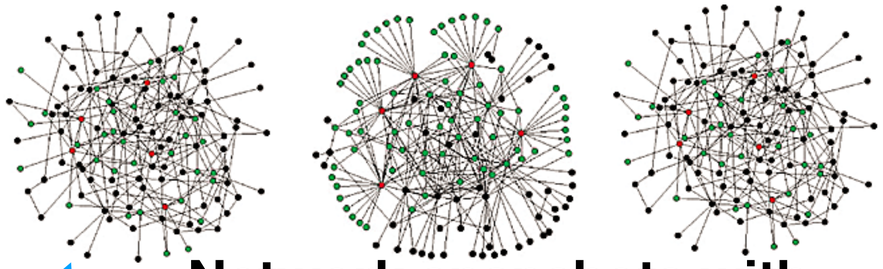
Monte Carlo Simulations

Candidate models

# How do we traditionally choose a model?




Network we want to model

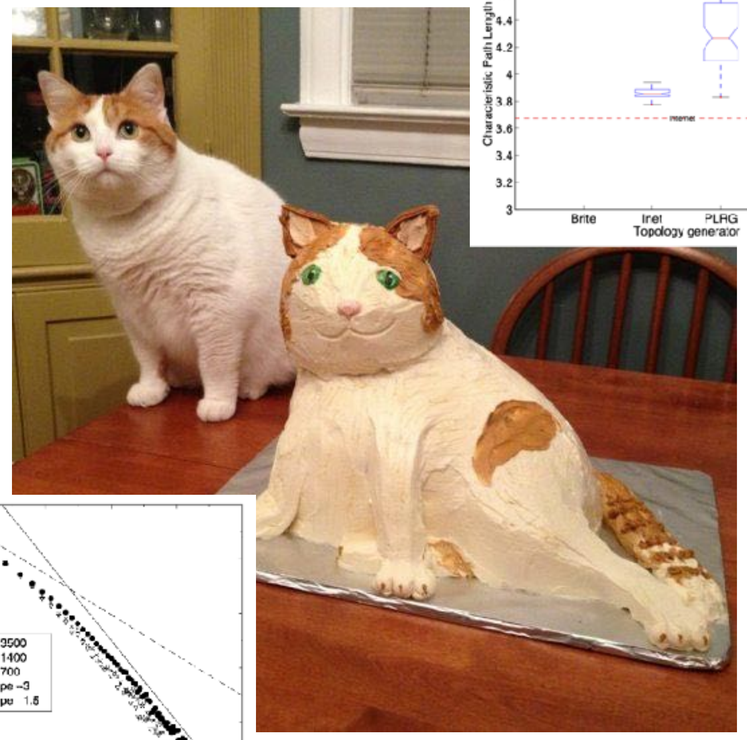


Network snapshots with same #nodes/links

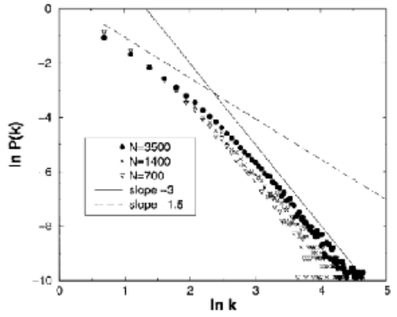
Monte Carlo Simulations

Candidate models

Comparison of stats with original network

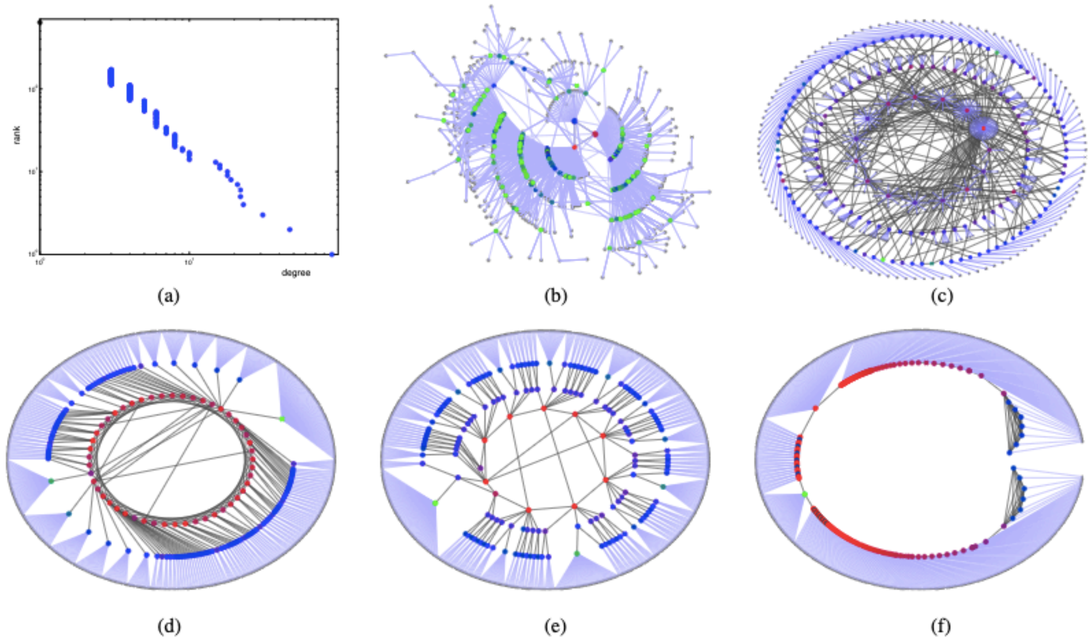




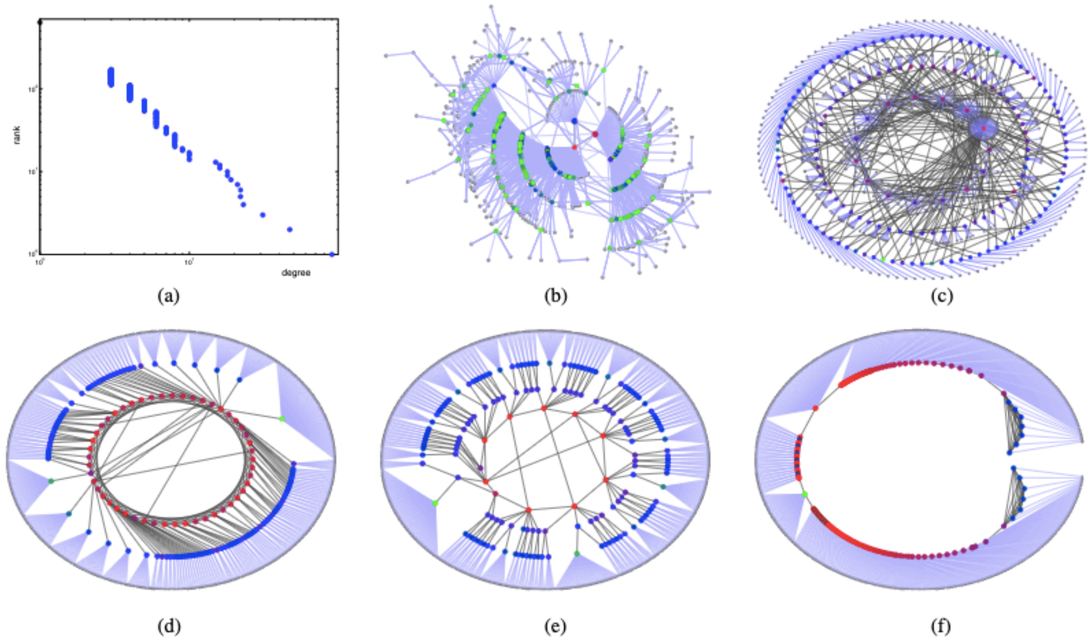
# Shortfalls with this approach

# Shortfalls with this approach




**1. Networks can have same statistics (e.g. degree distribution) but dramatically different properties**

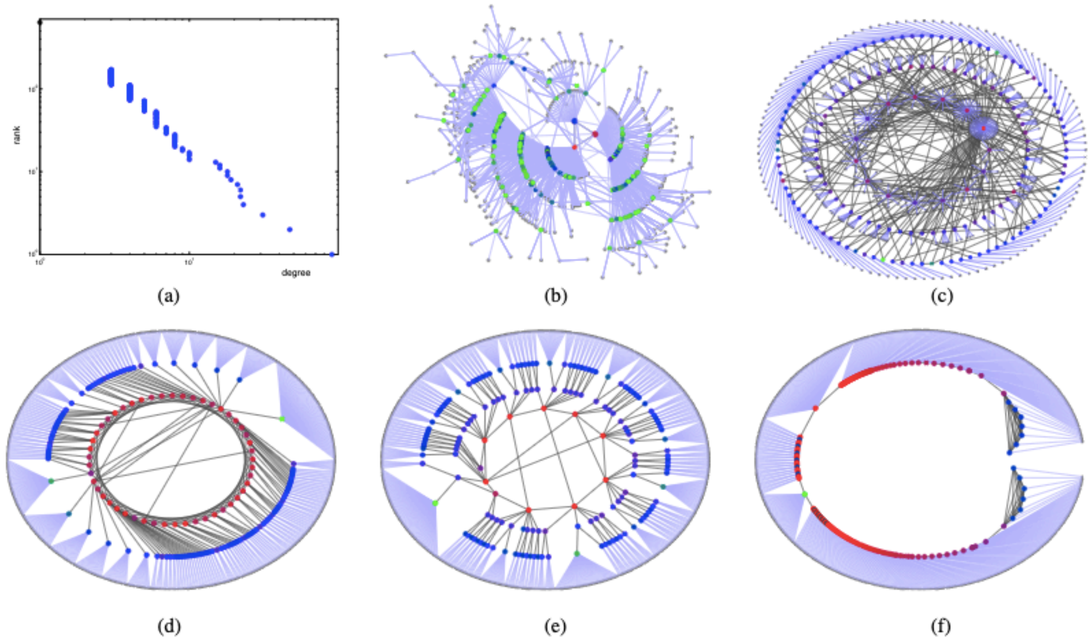
# Shortfalls with this approach




1. Networks can have same statistics (e.g. degree distribution) but dramatically different properties

2. Different models may perform better on different statistics

# Shortfalls with this approach

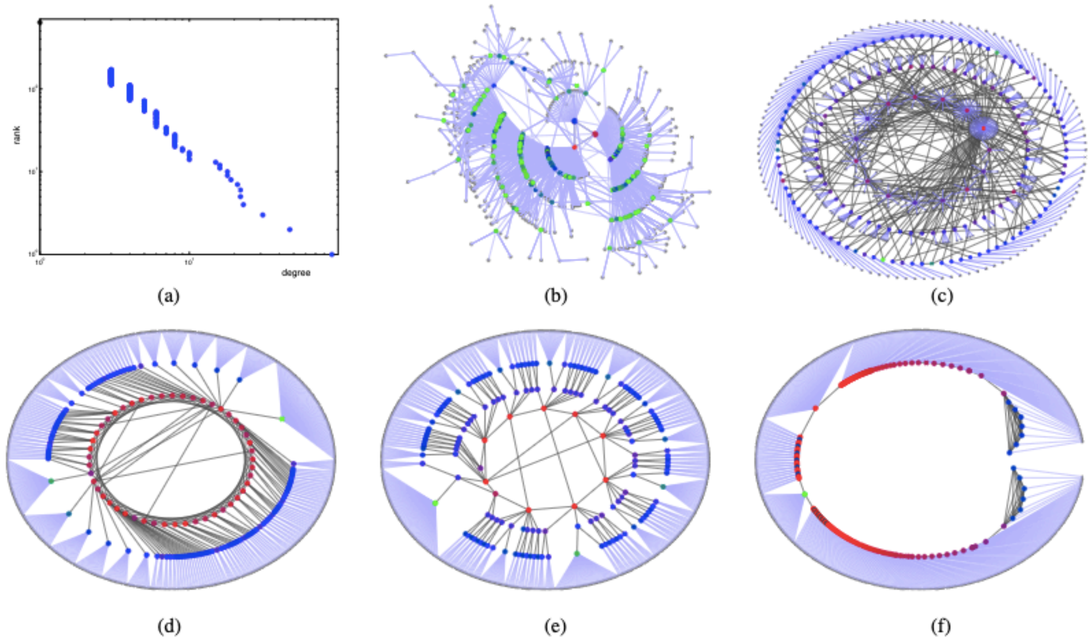



1. Networks can have same statistics (e.g. degree distribution) but dramatically different properties

2. Different models may perform better on different statistics

3. Doesn't capture any time-varying aspect of the network's growth

# Shortfalls with this approach



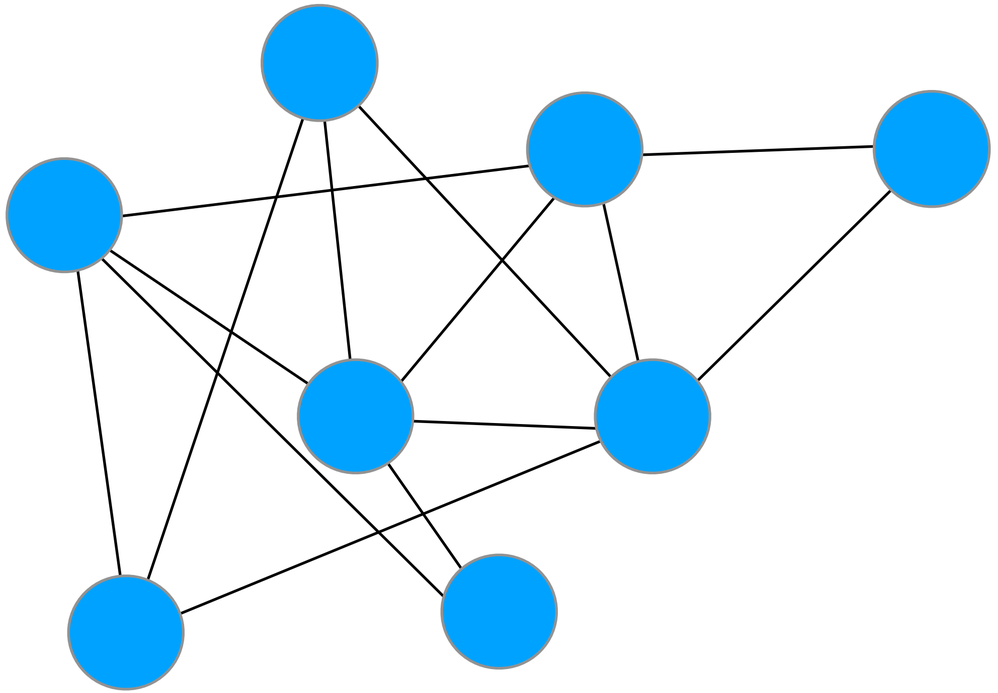

1. Networks can have same statistics (e.g. degree distribution) but dramatically different properties

2. Different models may perform better on different statistics

3. Doesn't capture any time-varying aspect of the network's growth

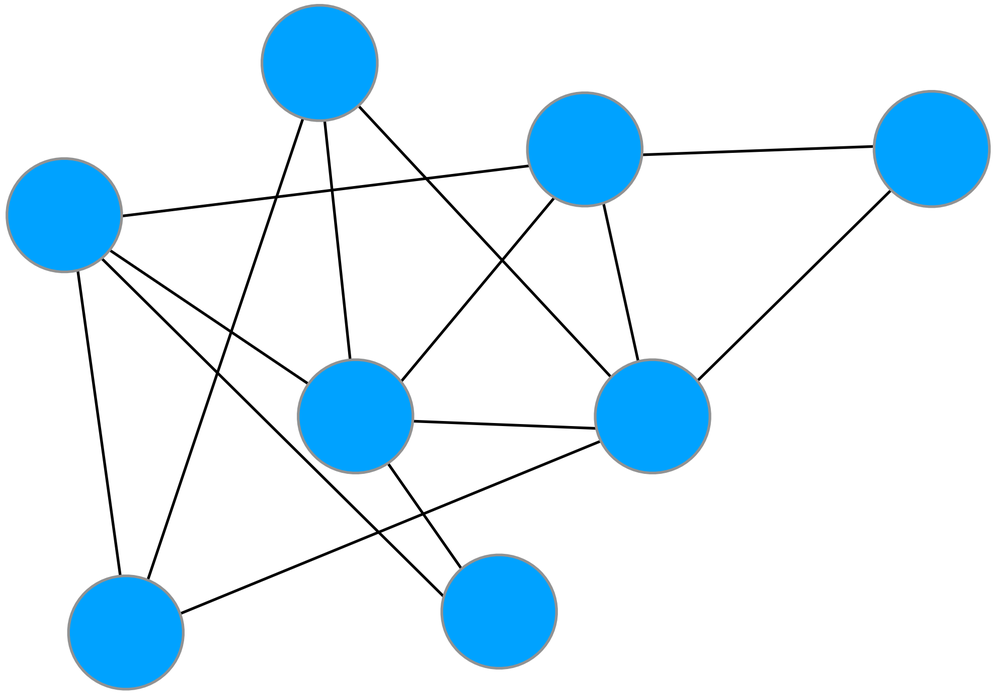
Hmm… What about with more information than just a snapshot?

# With temporal information of link arrival times, we can do better!

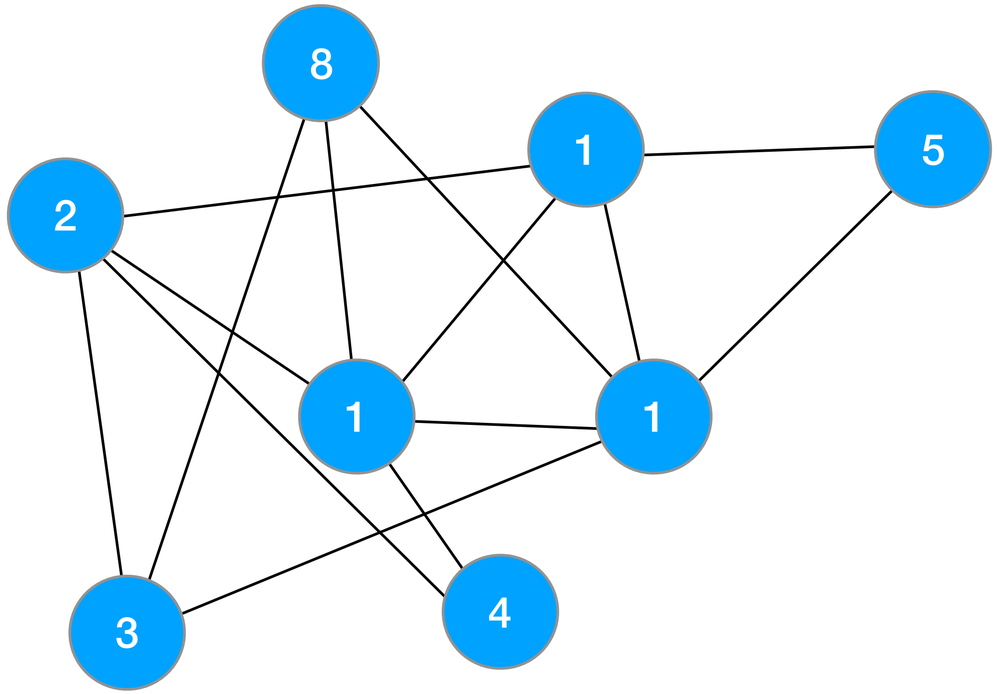


**vs**

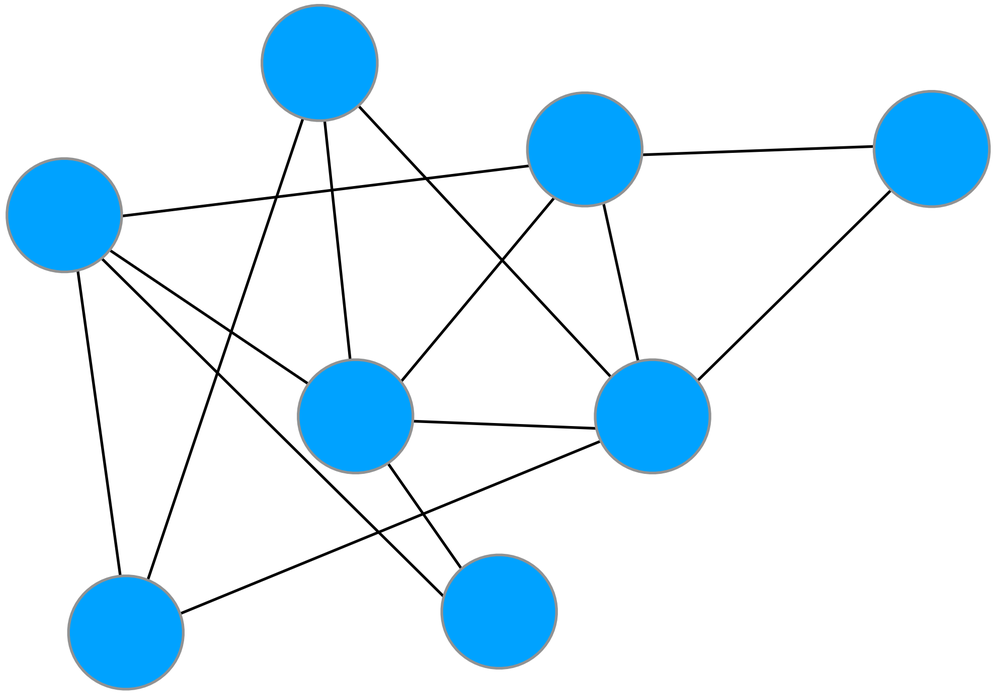
# With temporal information of link arrival times, we can do better!



vs

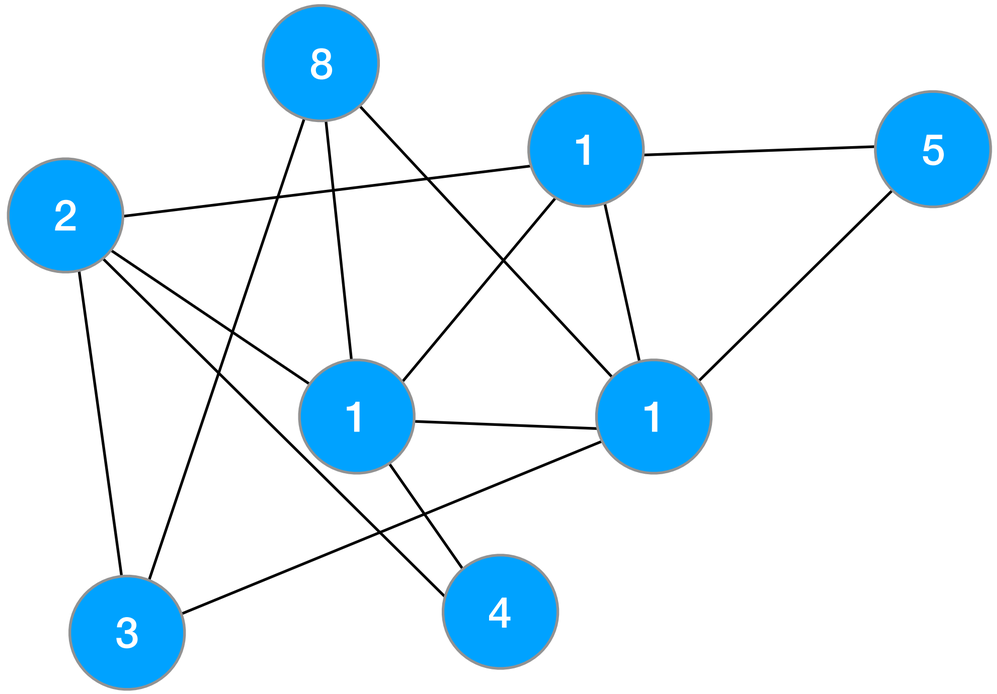


# With temporal information of link arrival times, we can do better!



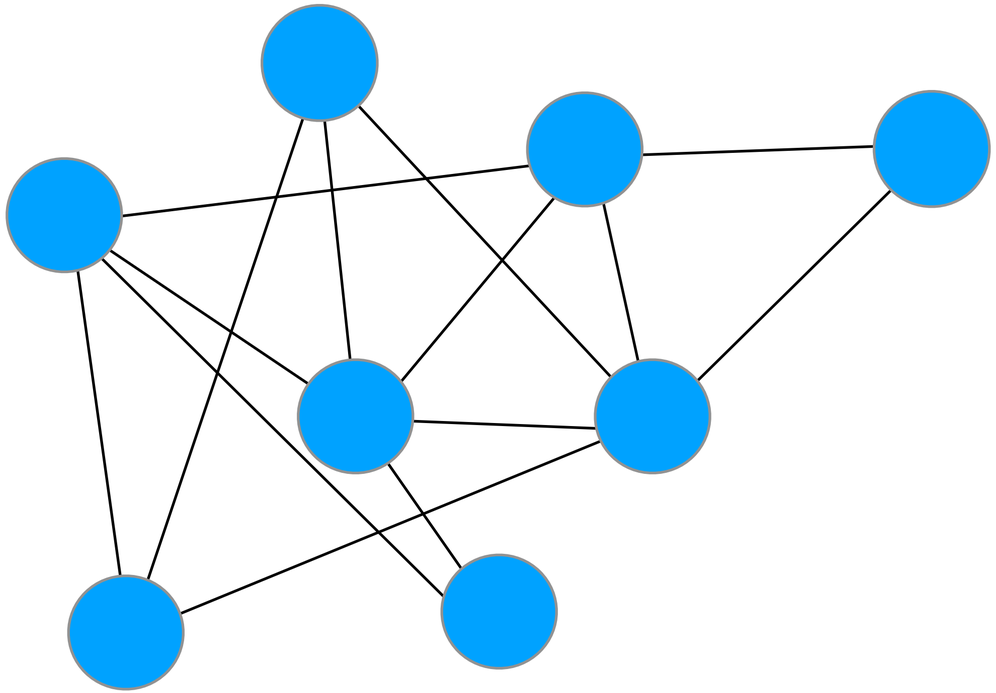
**n=1 graph observations**

**vs**

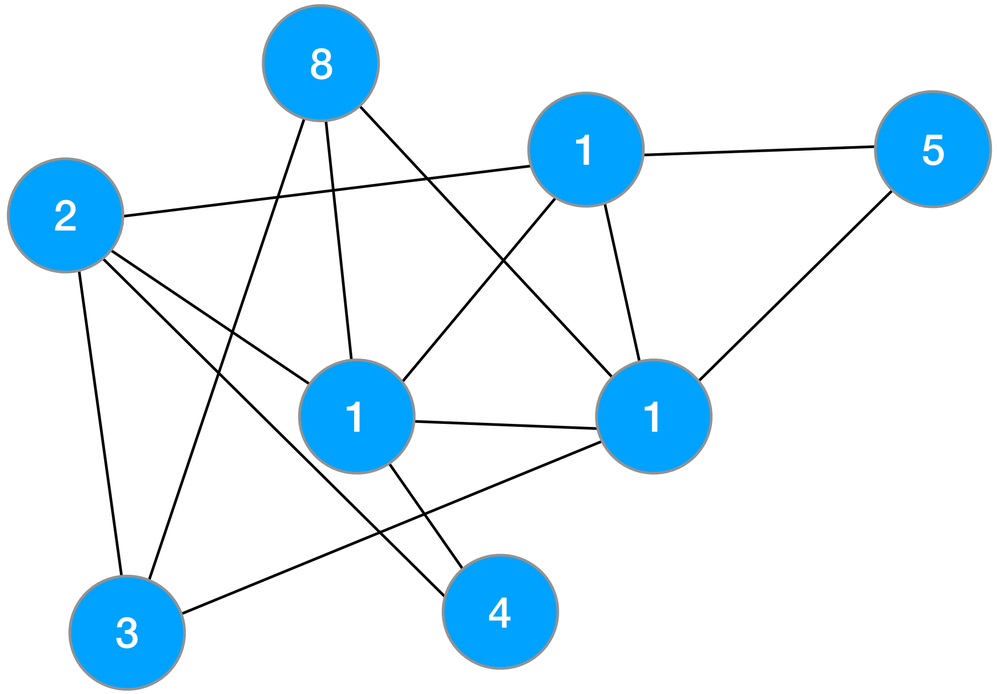



**n=8 observations of graph growth**

# With temporal information of link arrival times, we can do better!



vs



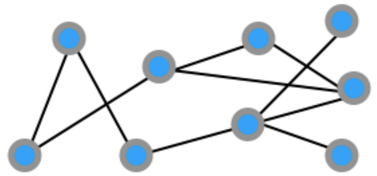

**n=1 graph observations**

**n=8 observations of graph growth**

**=> Can calculate precise likelihood of model, see R.Clegg et al: Likelihood based assessment of dynamic networks (2015)**

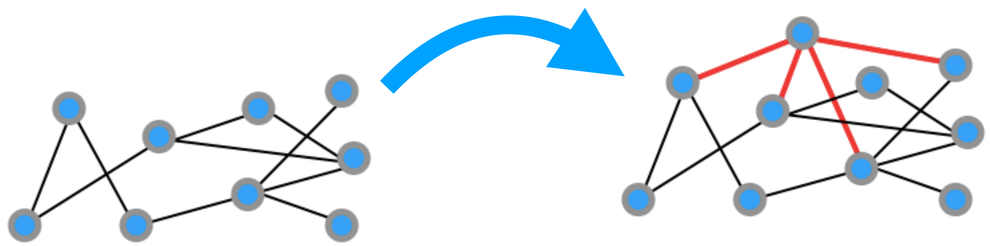
# What does such a model look like?

**For any new node joining the network, or existing node choosing to make new connections, node i is chosen as a neighbour with probability:**



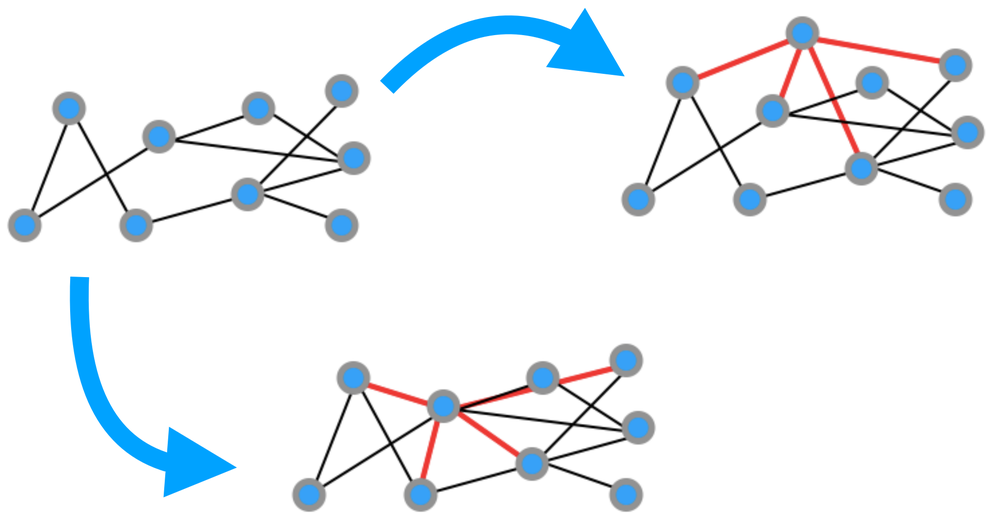
# What does such a model look like?

**For any new node joining the network, or existing node choosing to make new connections, node i is chosen as a neighbour with probability:**



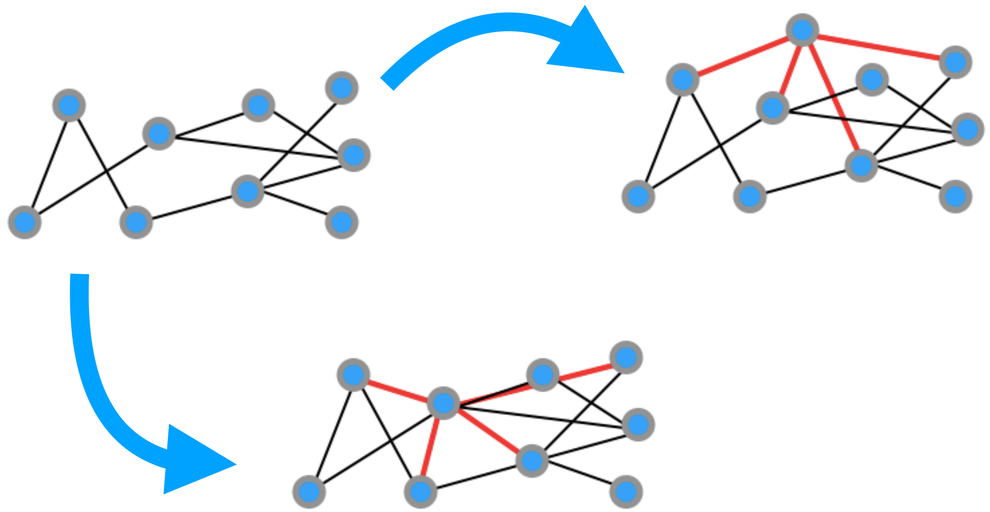
# What does such a model look like?

**For any new node joining the network, or existing node choosing to make new connections, node i is chosen as a neighbour with probability:**



# What does such a model look like?

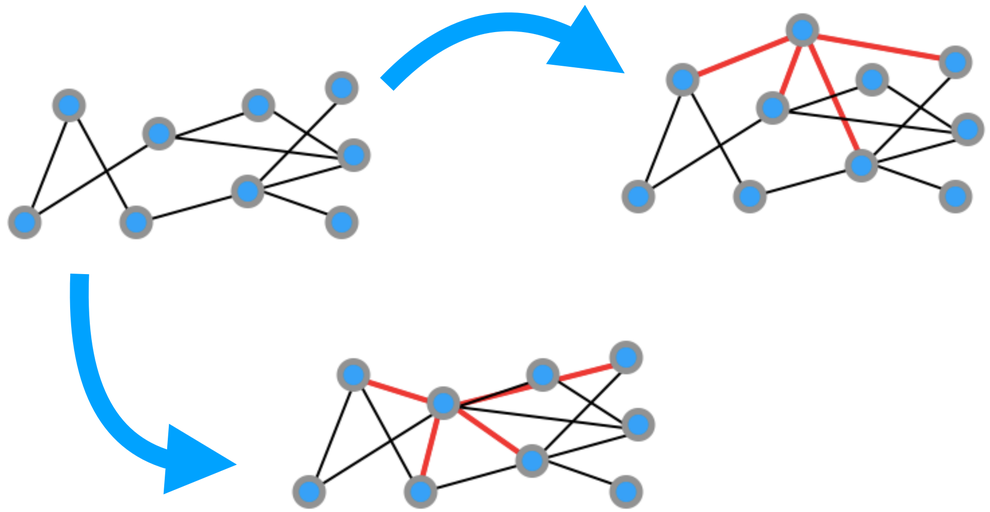
**For any new node joining the network, or existing node choosing to make new connections, node i is chosen as a neighbour with probability:**



$$p_i = \beta_1(t)p_i^{(1)} + \ldots + \beta_l(t)p_i^{(l)}$$

# What does such a model look like?

**For any new node joining the network, or existing node choosing to make new connections, node i is chosen as a neighbour with probability:**

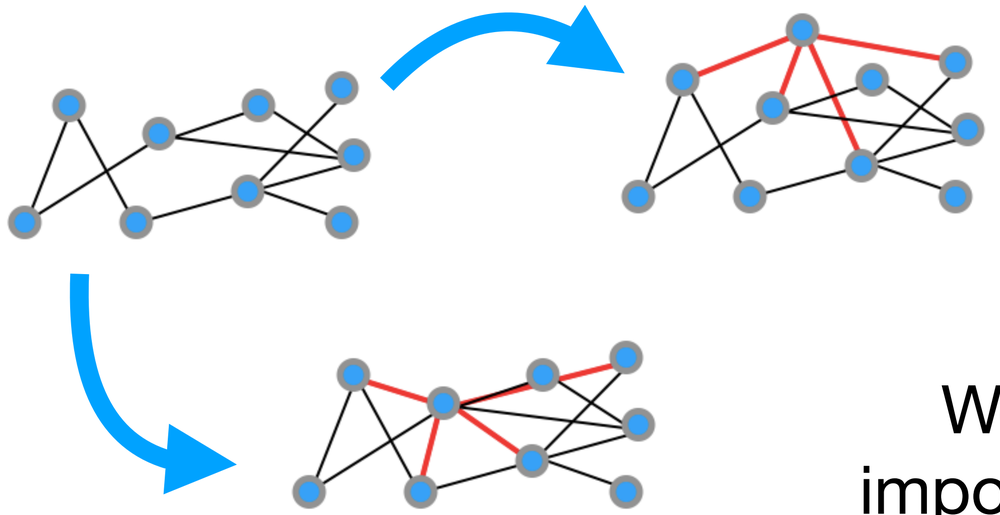


$$p_i = \beta_1(t)p_i^{(1)} + \ldots + \beta_l(t)p_i^{(l)}$$

Sum is over probabilities according to different models, e.g. Preferential Attachment/Triangle Closure

# What does such a model look like?

**For any new node joining the network, or existing node choosing to make new connections, node i is chosen as a neighbour with probability:**



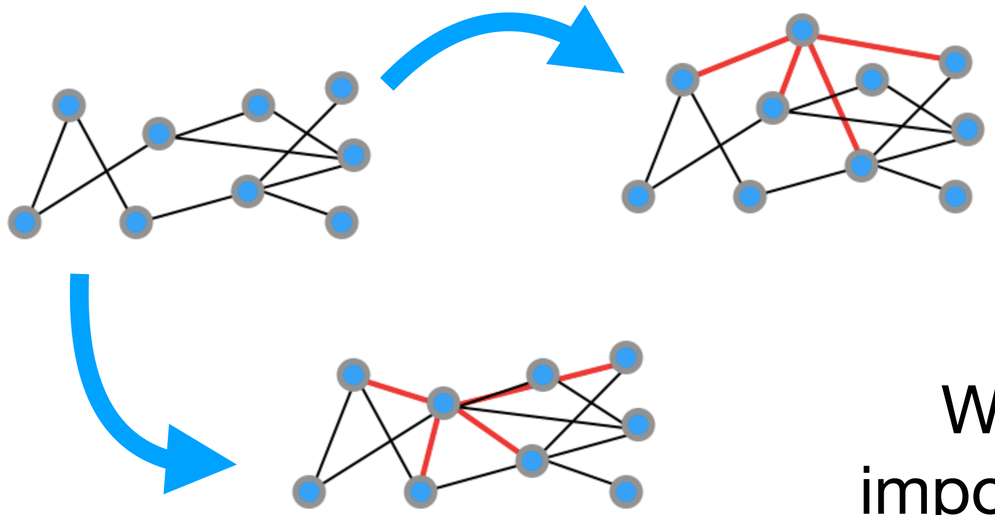
$$p_i = \beta_1(t)p_i^{(1)} + \ldots + \beta_l(t)p_i^{(l)}$$

Sum is over probabilities according to different models, e.g. Preferential Attachment/Triangle Closure

Weights show importance of each model

# What does such a model look like?

**For any new node joining the network, or existing node choosing to make new connections, node i is chosen as a neighbour with probability:**



$$p_i = \beta_1(t)p_i^{(1)} + \ldots + \beta_l(t)p_i^{(l)}$$

Sum is over probabilities according to different models, e.g. Preferential Attachment/Triangle Closure

Weights show importance of each model

This importance may change over time

# Artificial data example

$$p_i(t) \propto \begin{cases} k_i^{\alpha} & t \leq T \\ k_i^{\beta} & t > T \end{cases}$$

**Preferential attachment with a strength (exponent) that abruptly changes at time T**

**Model without changepoint found in Krapivsky et al: Connectivity of Growing Random Networks (2000)**

# Artificial data example

$$p_i(t) \propto \begin{cases} k_i^{\alpha} & t \leq T \\ k_i^{\beta} & t > T \end{cases}$$

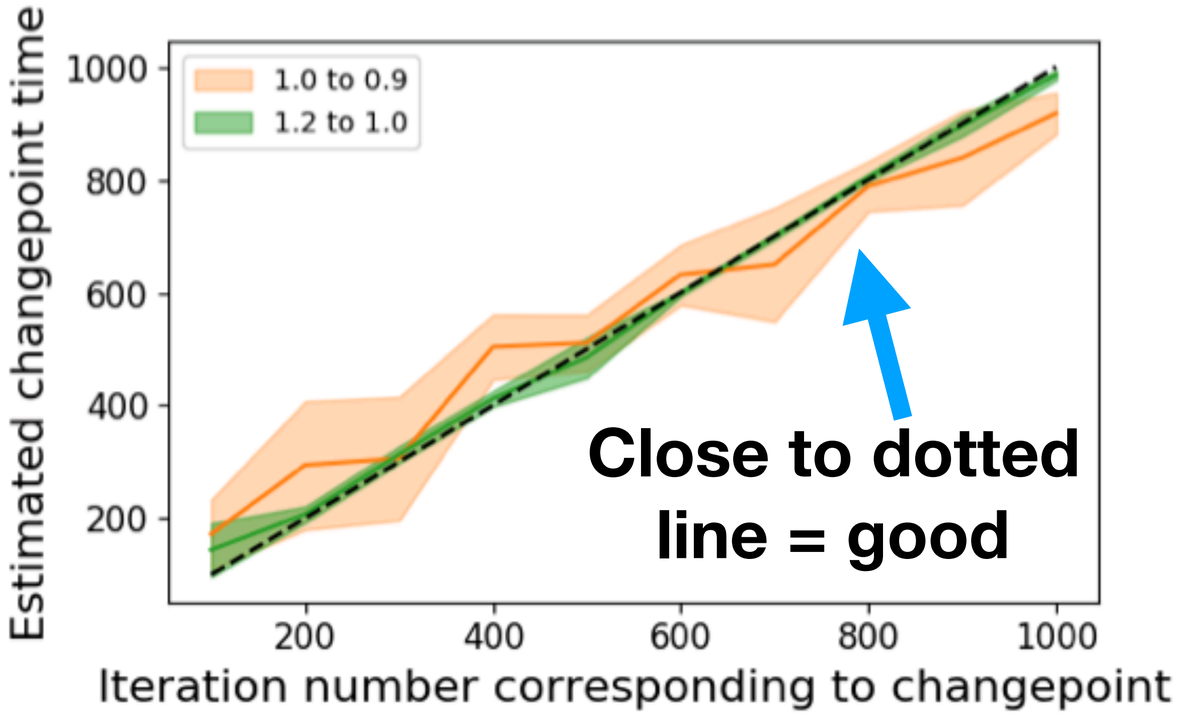
**Preferential attachment with a strength (exponent) that abruptly changes at time T**

**Experiment: Create artificial data set with different $T$ and try to retrieve it**

**Model without changepoint found in Krapivsky et al: Connectivity of Growing Random Networks (2000)**

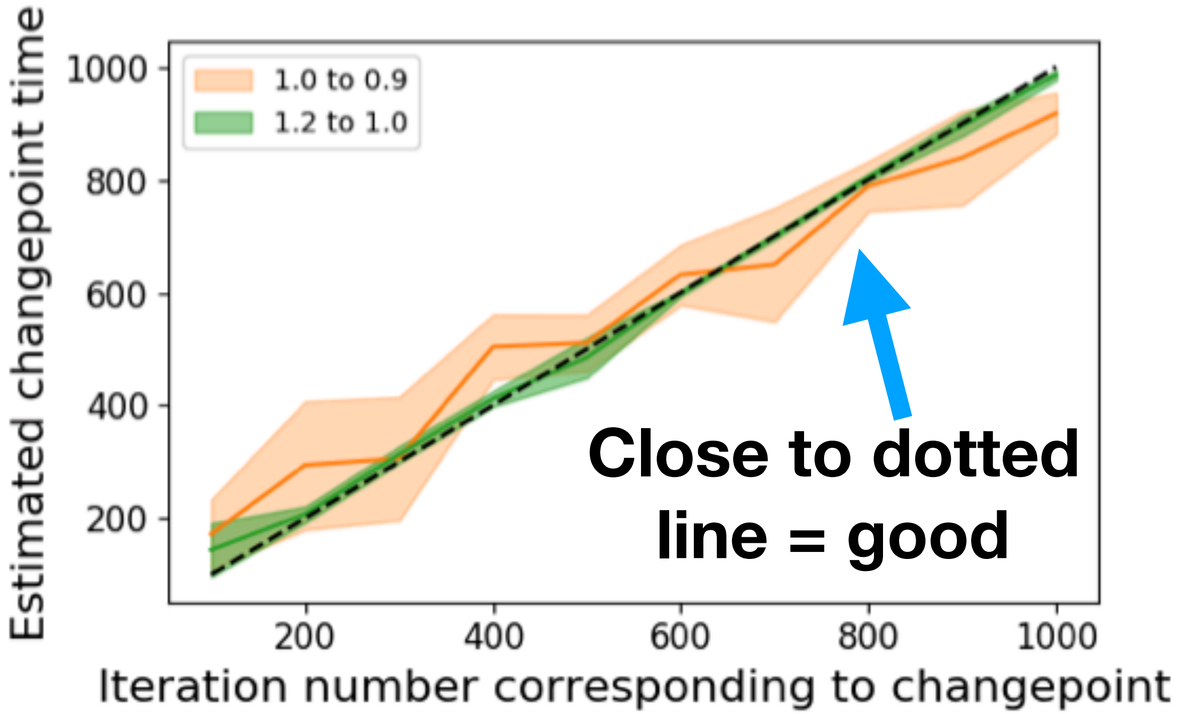
# Example: Nonlinear Preferential Attachment

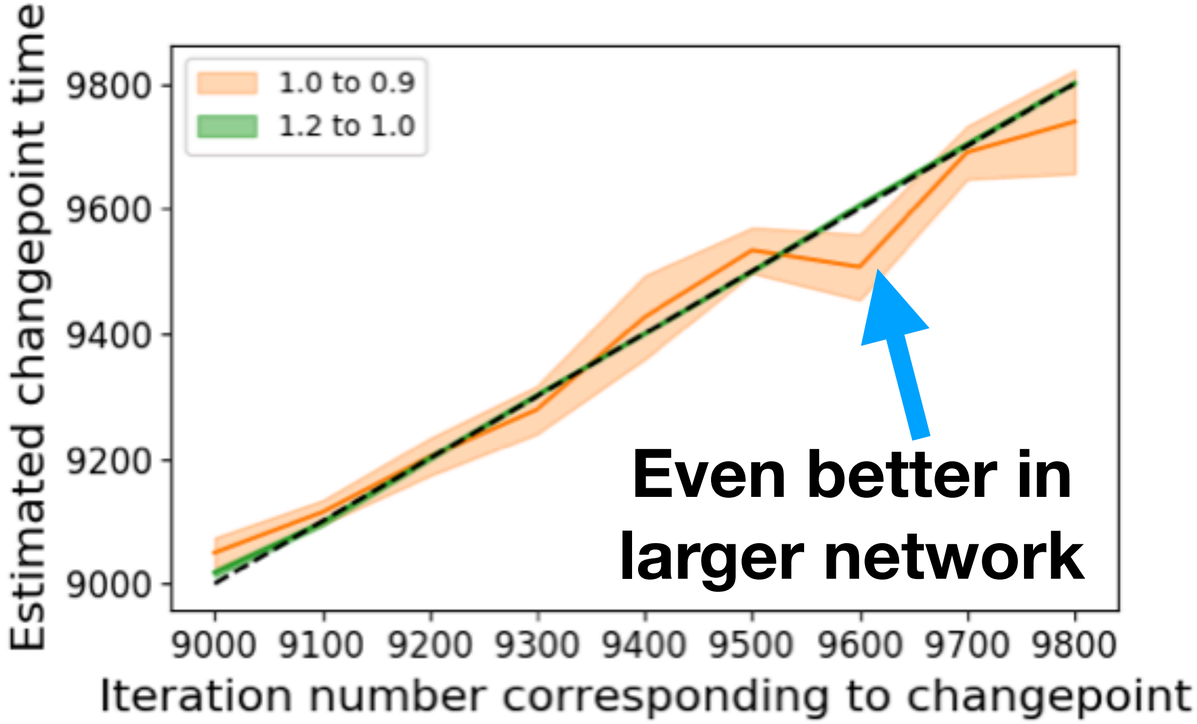
# Example: Nonlinear Preferential Attachment




**1,000 node network**

# Example: Nonlinear Preferential Attachment



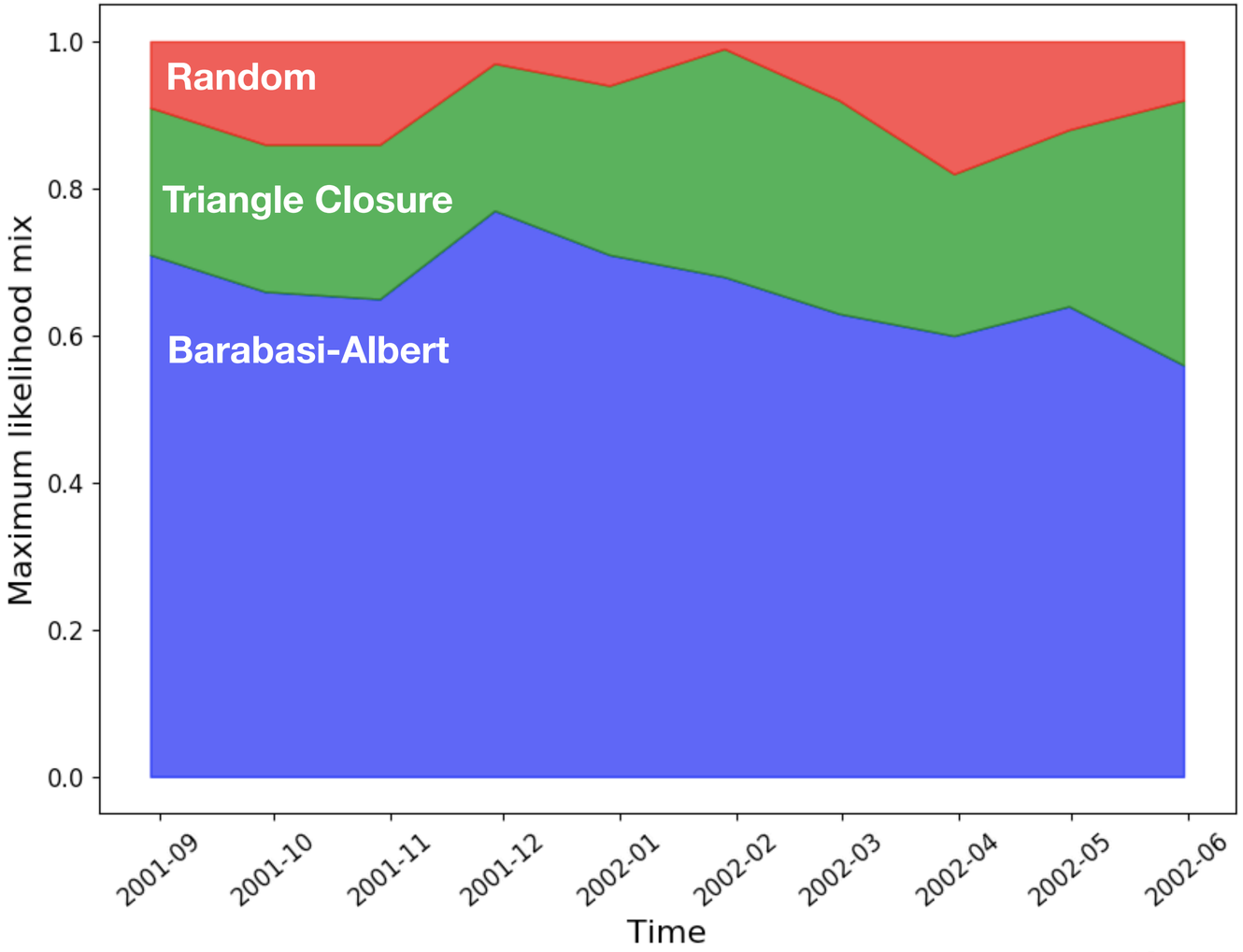




**1,000 node network**

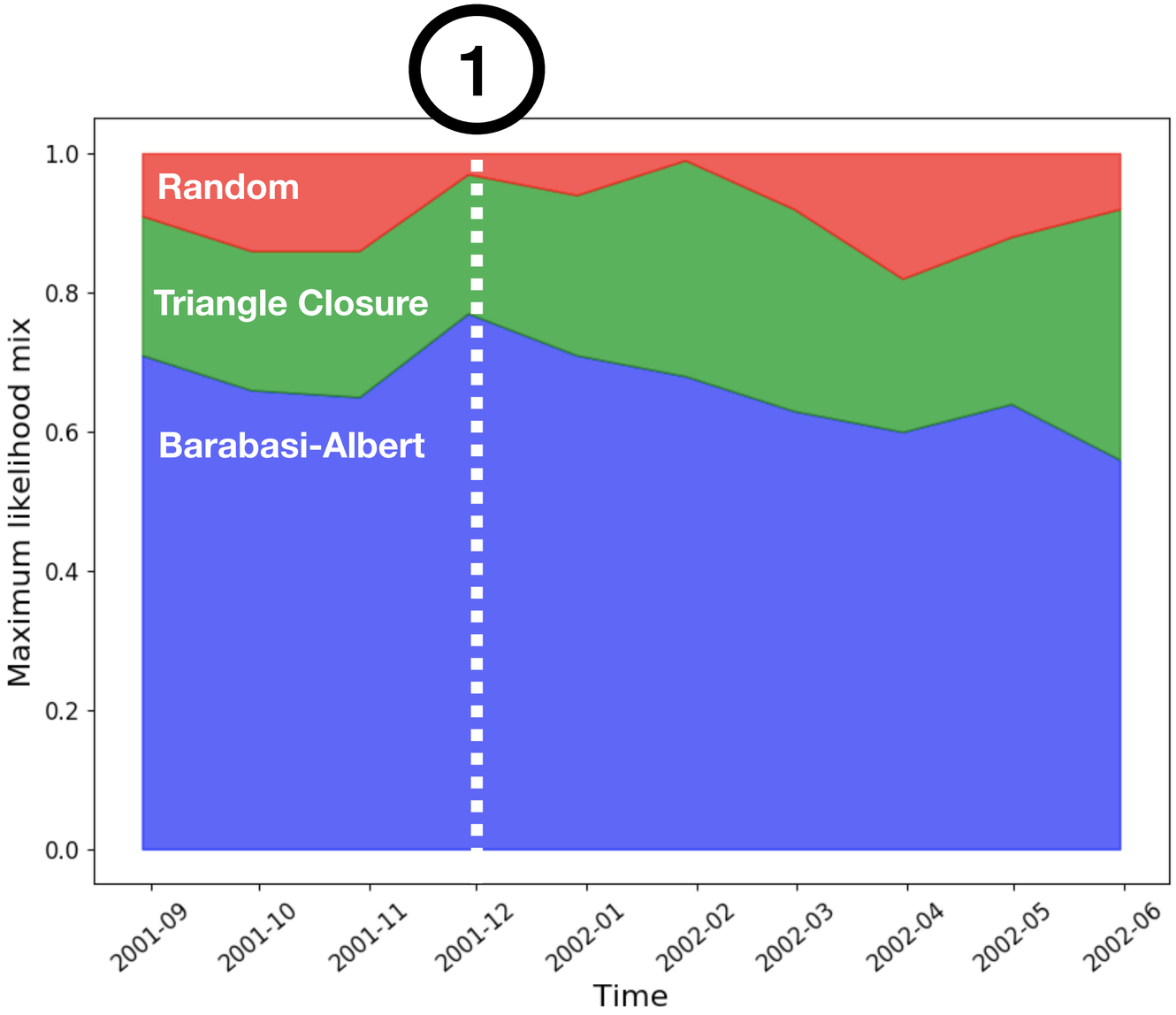
**10,000 node network**

# Enron Revisited

# Enron Revisited

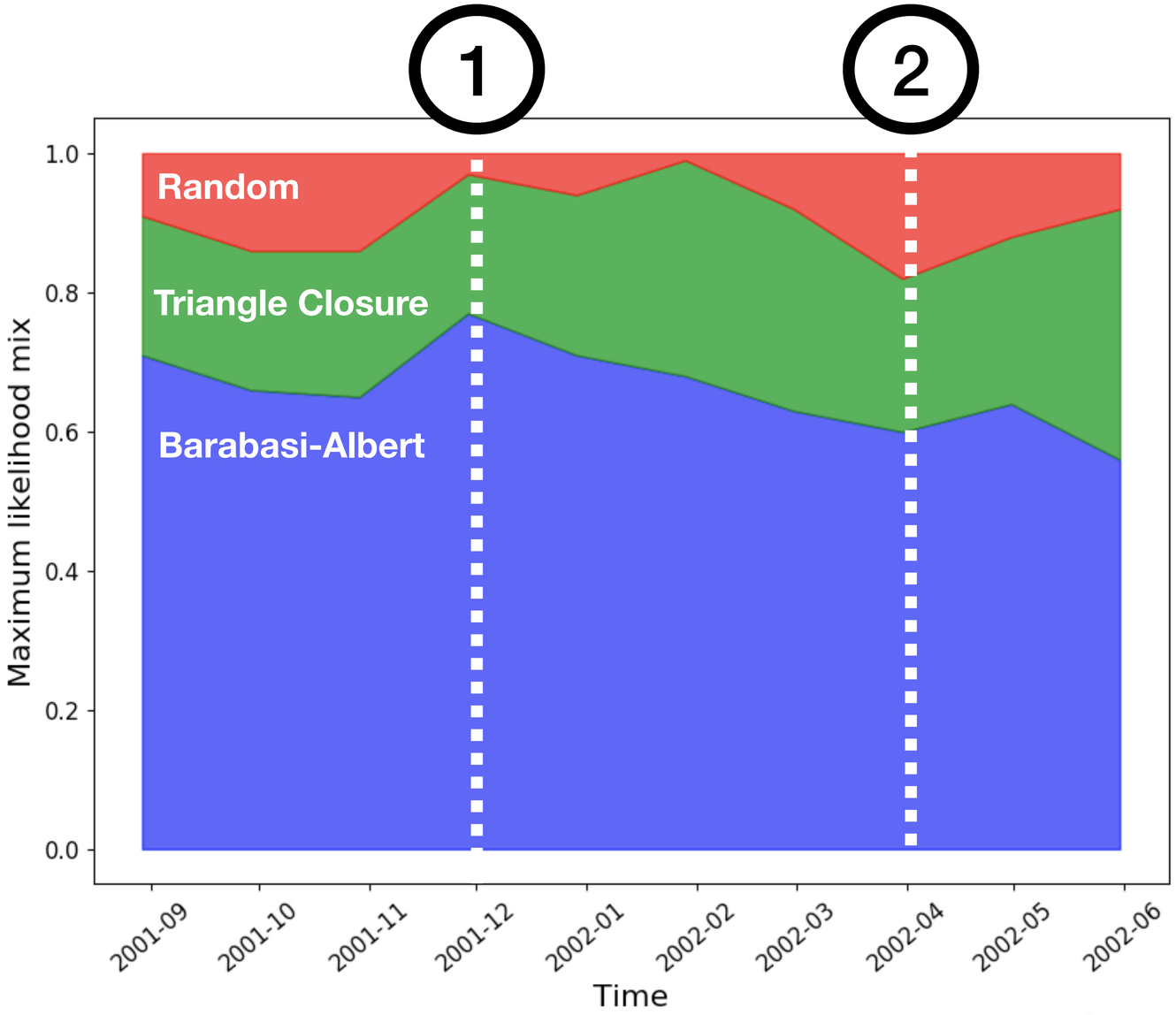


# Enron Revisited




1 **Dec 2, 2001: Enron goes bankrupt, thousands of workers laid off**

# Enron Revisited




1 **Dec 2, 2001: Enron goes bankrupt, thousands of workers laid off**

2 **April 9, 2001: Top Enron auditor pleads guilty to obstruction for ordering staff to destroy documents**

# Takeaways

# Takeaways

- Often a **mixture of mechanisms** better describes a network's growth rather than a single one.

# Takeaways

- Often a **mixture of mechanisms** better describes a network's growth rather than a single one.
- This mixture may **change over time**, which may tell us about a network's response to events as well as longer term trends.

# Takeaways

- Often a **mixture of mechanisms** better describes a network's growth rather than a single one.
- This mixture may **change over time**, which may tell us about a network's response to events as well as longer term trends.
- Framework for combining these mechanisms gives us a new way of analysing growing networks

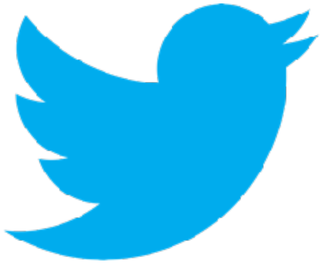

**Framework for Evolving Topology Analysis**

https://github.com/narnolddd/FETA3

# Thanks for listening!
# Questions?

**github.com/narnolddd**

**n.a.arnold@qmul.ac.uk**



**@narnolddd**