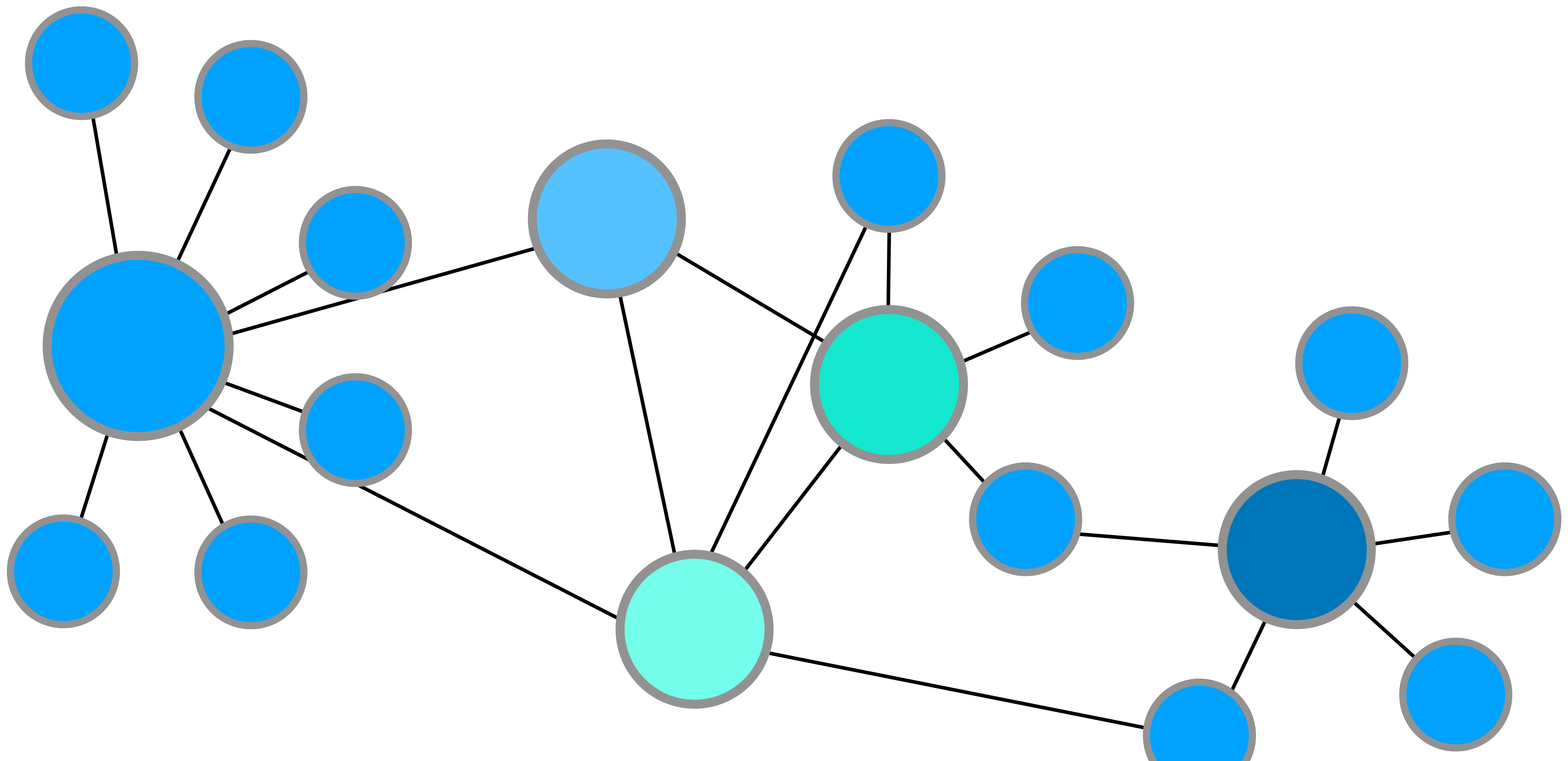
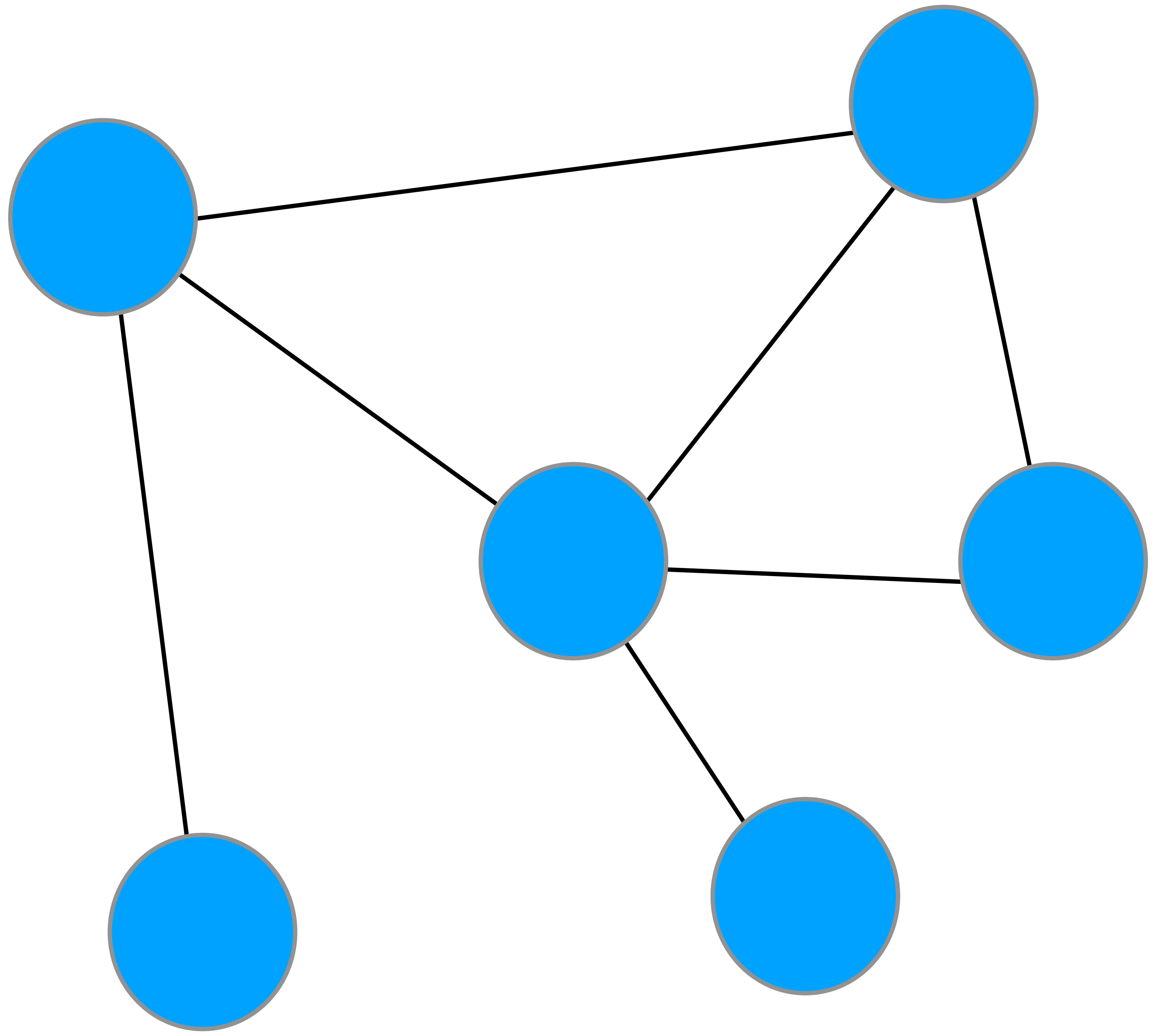
Uncovering network evolution mechanisms using temporal data

Naomi Arnold, Raul Mondragon, Richard Clegg



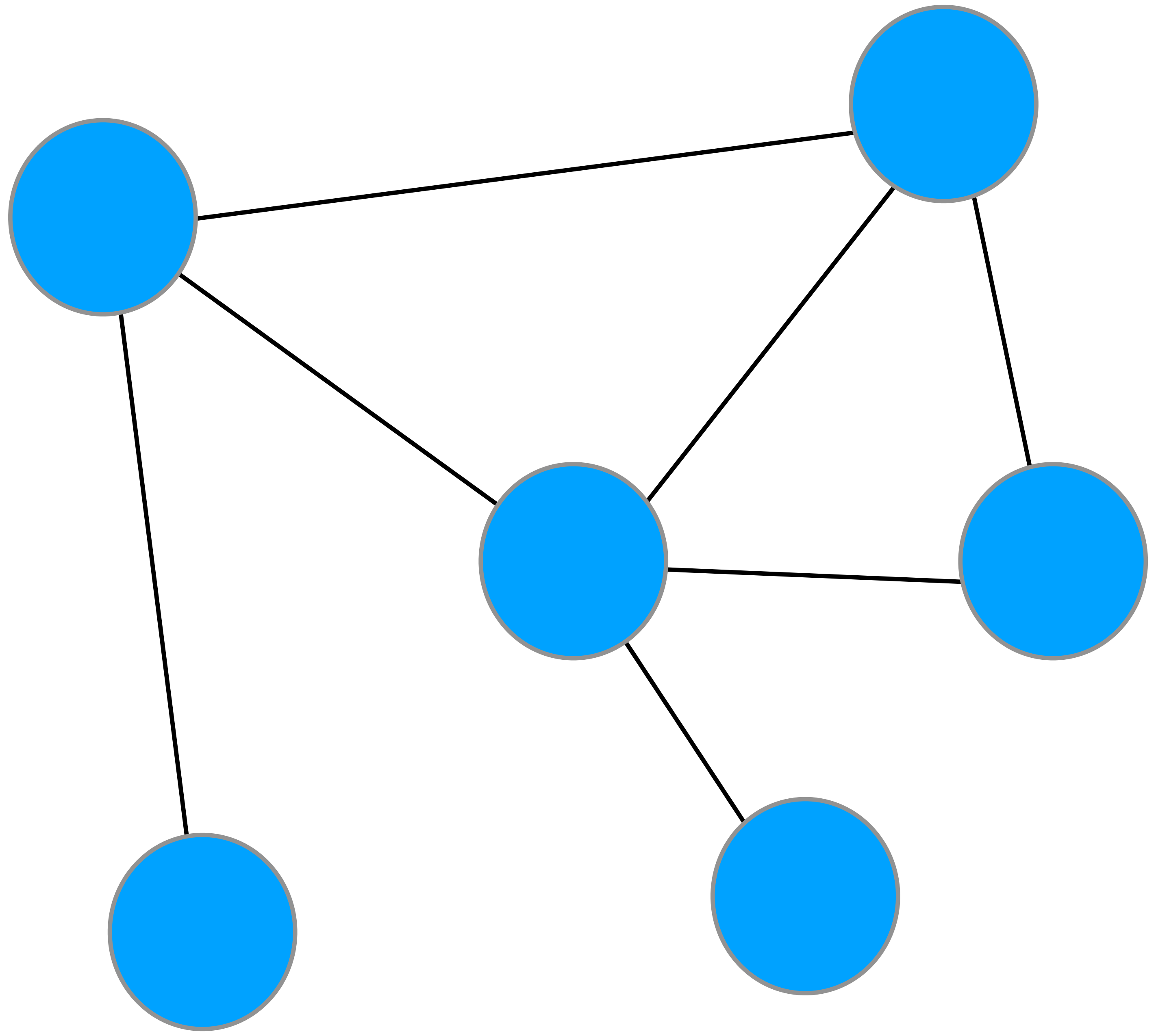
Models for network topology

Models for network topology



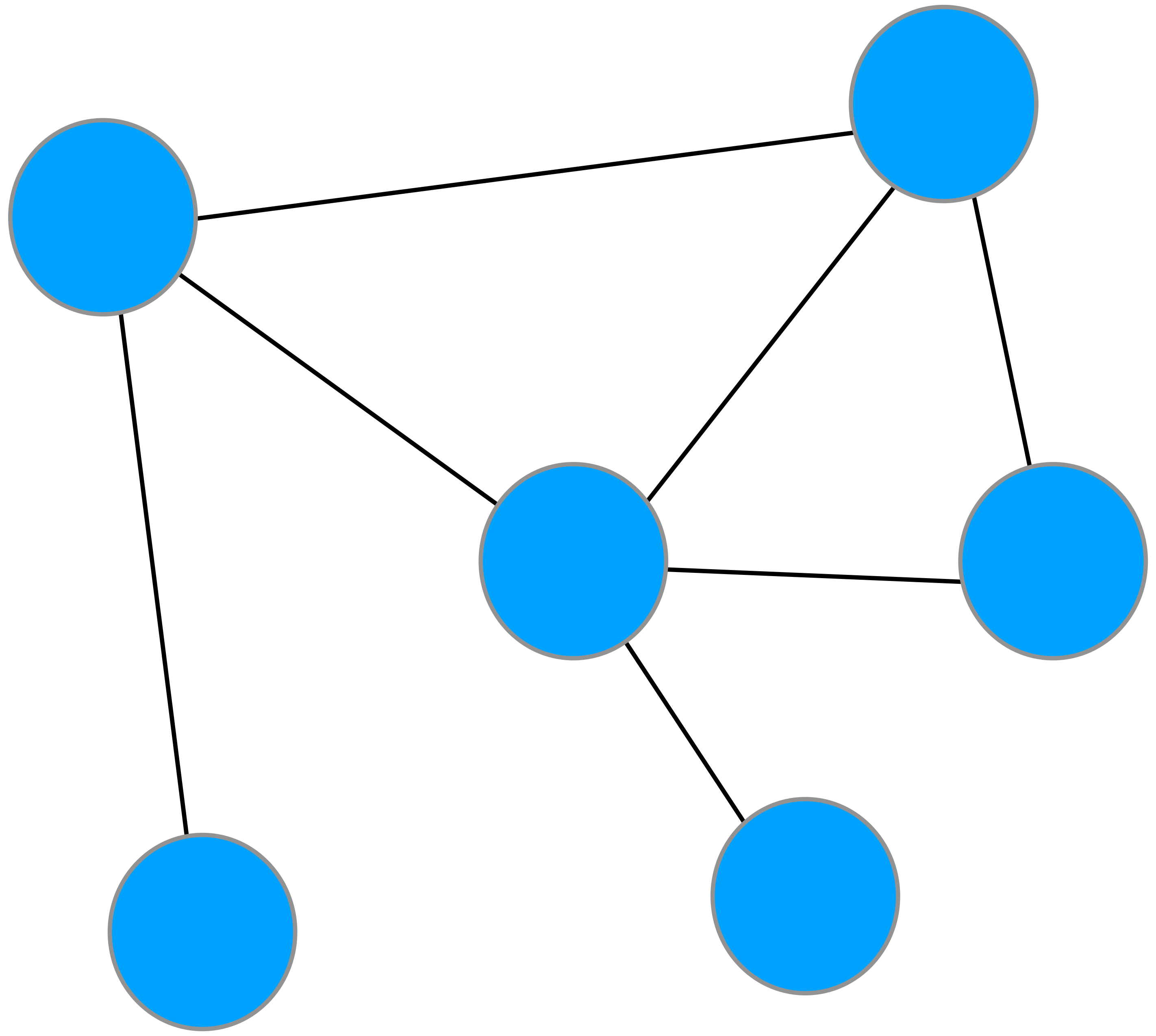
Static Models

Models for network topology

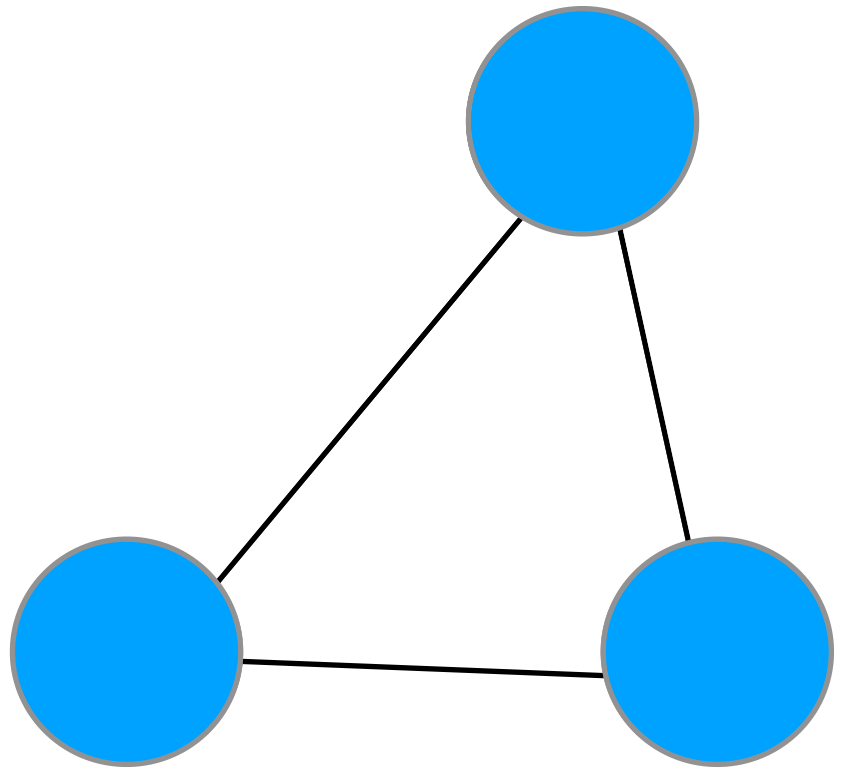


Static Models **Dynamic Models**

Models for network topology

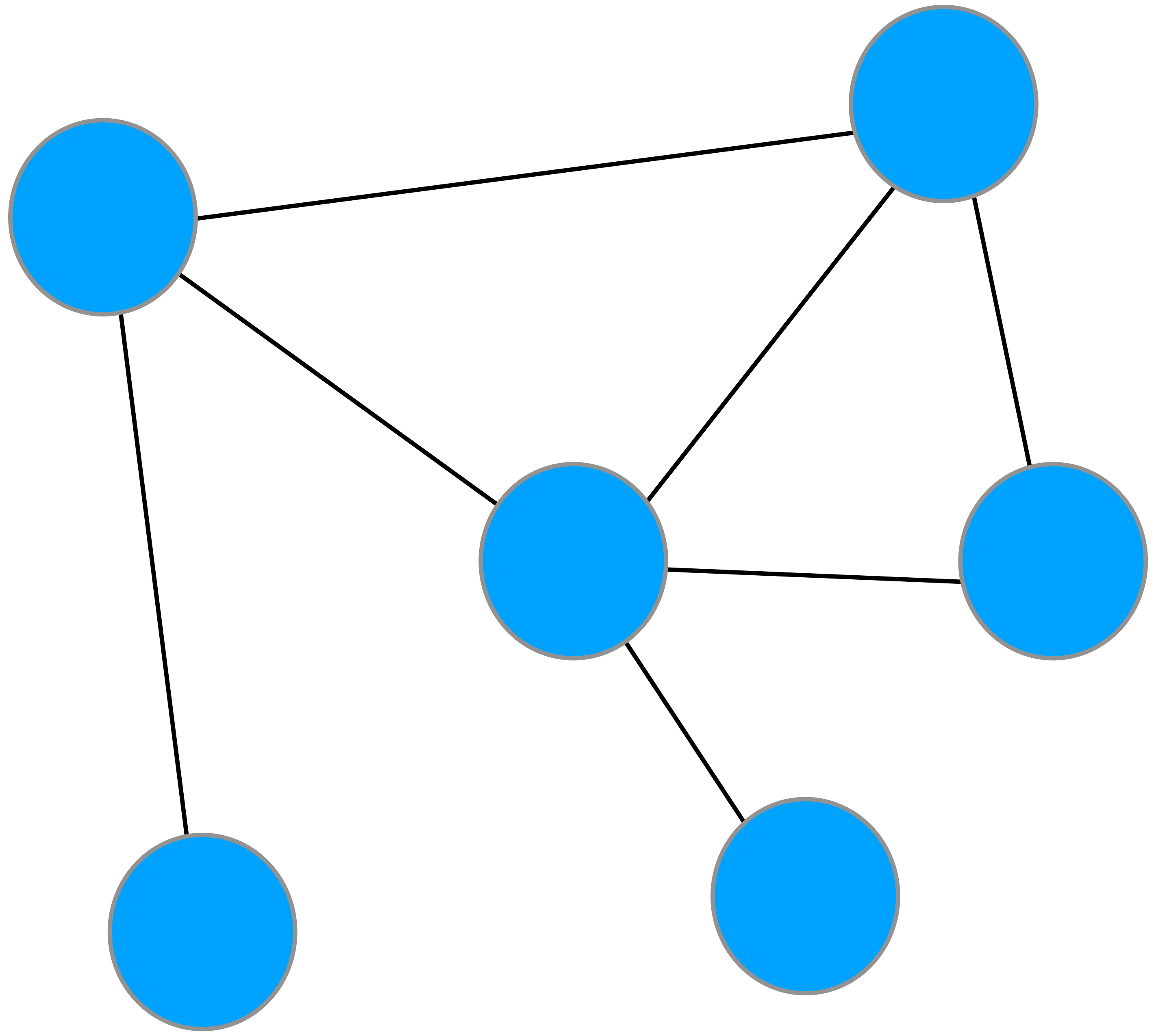


Static Models

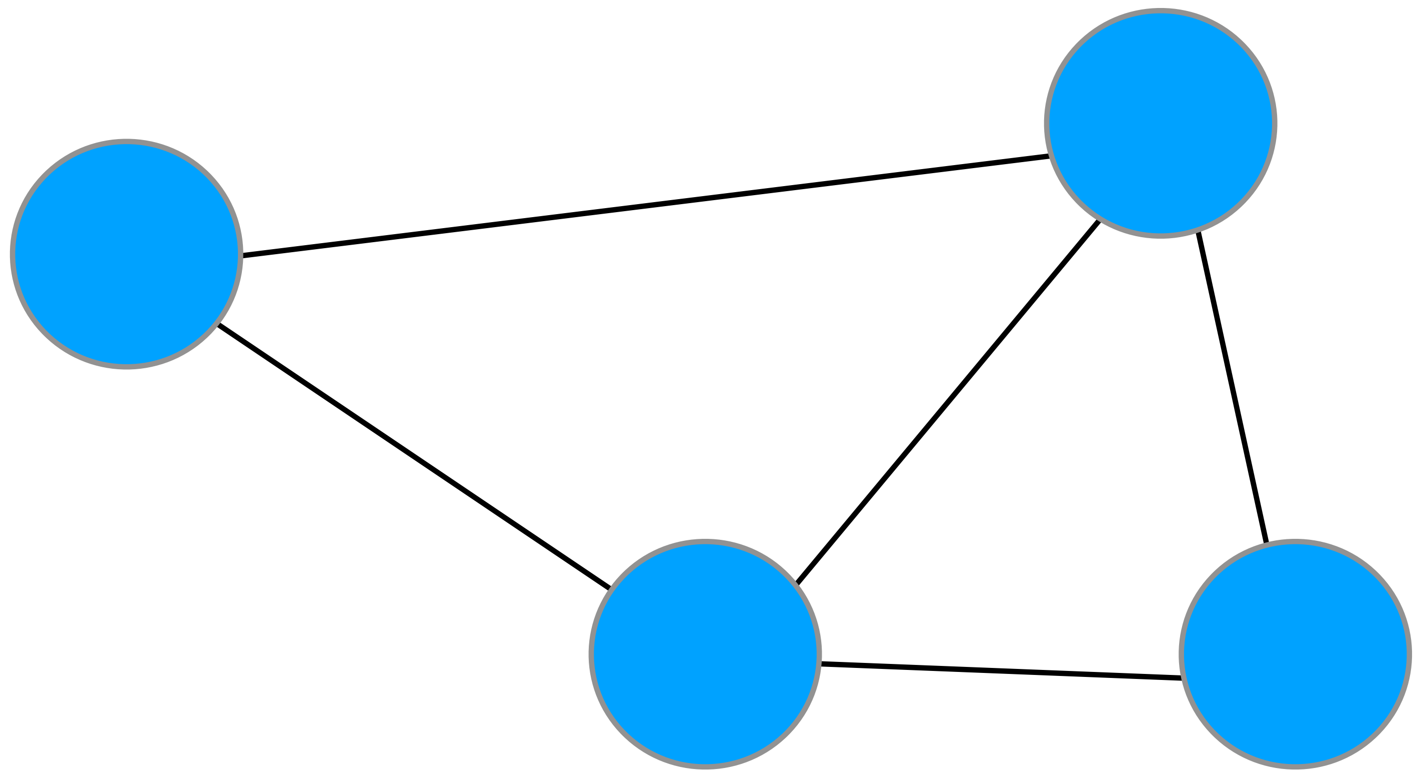


Dynamic Models

Models for network topology

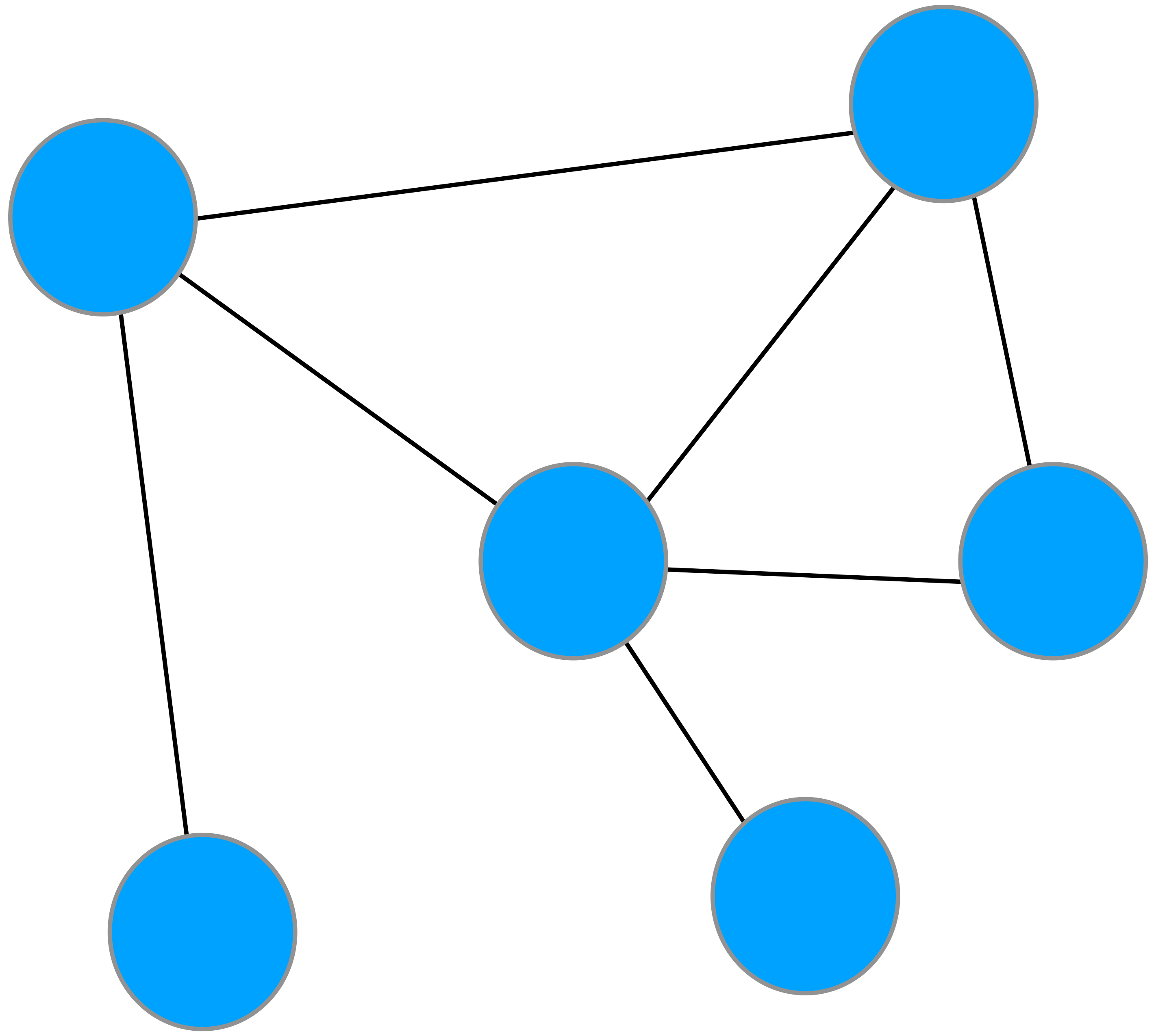


Static Models

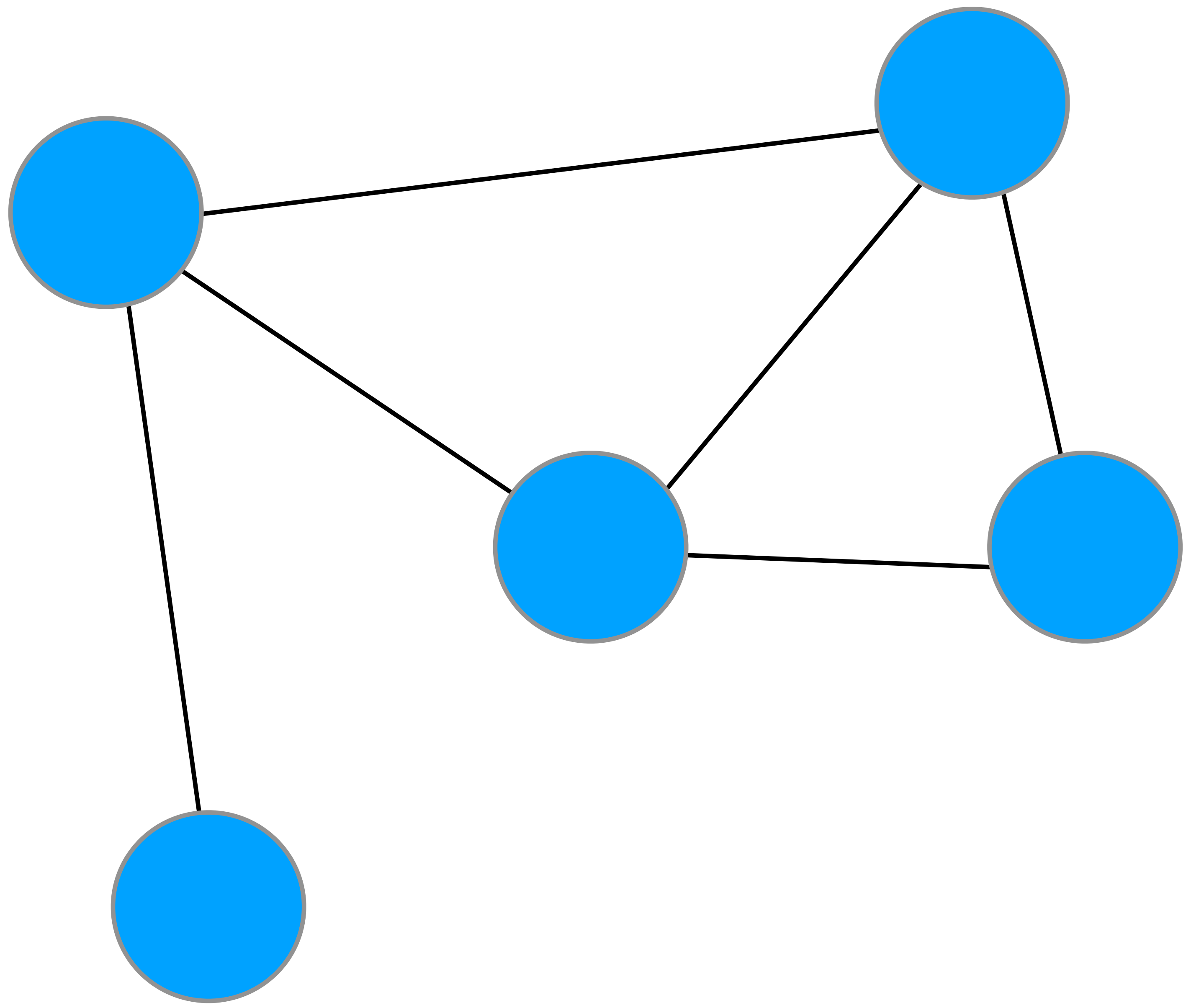


Dynamic Models

Models for network topology

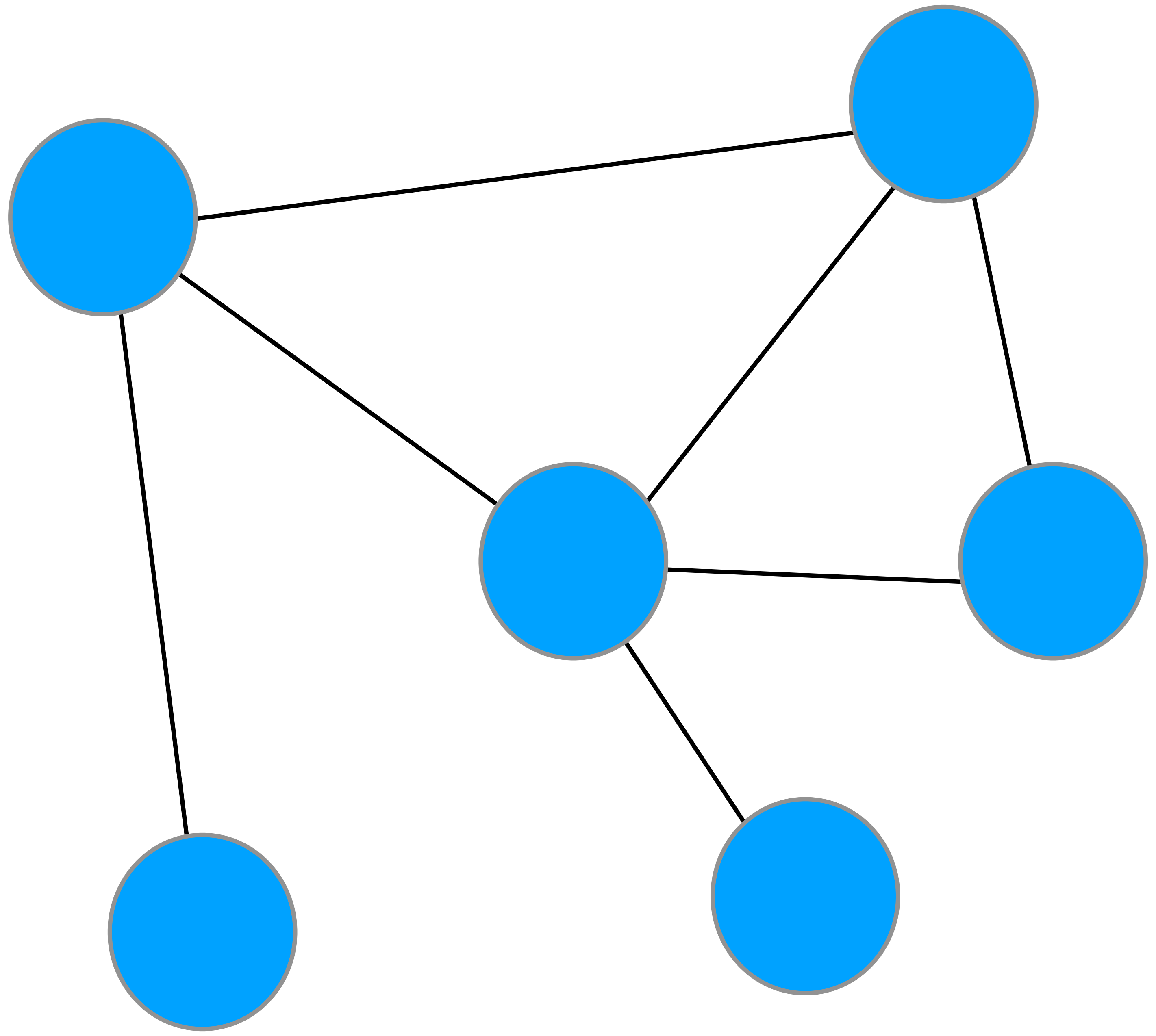


Static Models

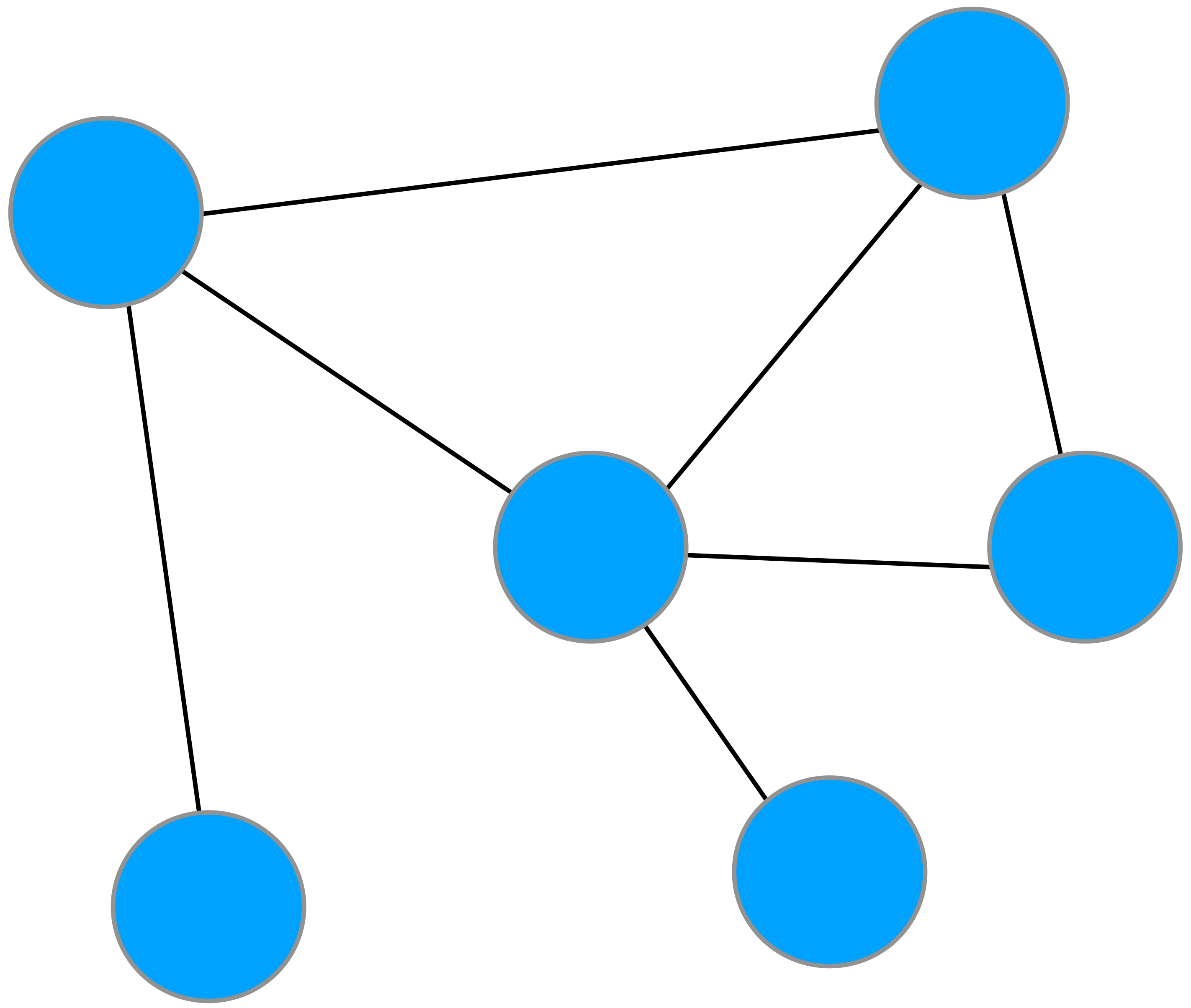


Dynamic Models

Models for network topology

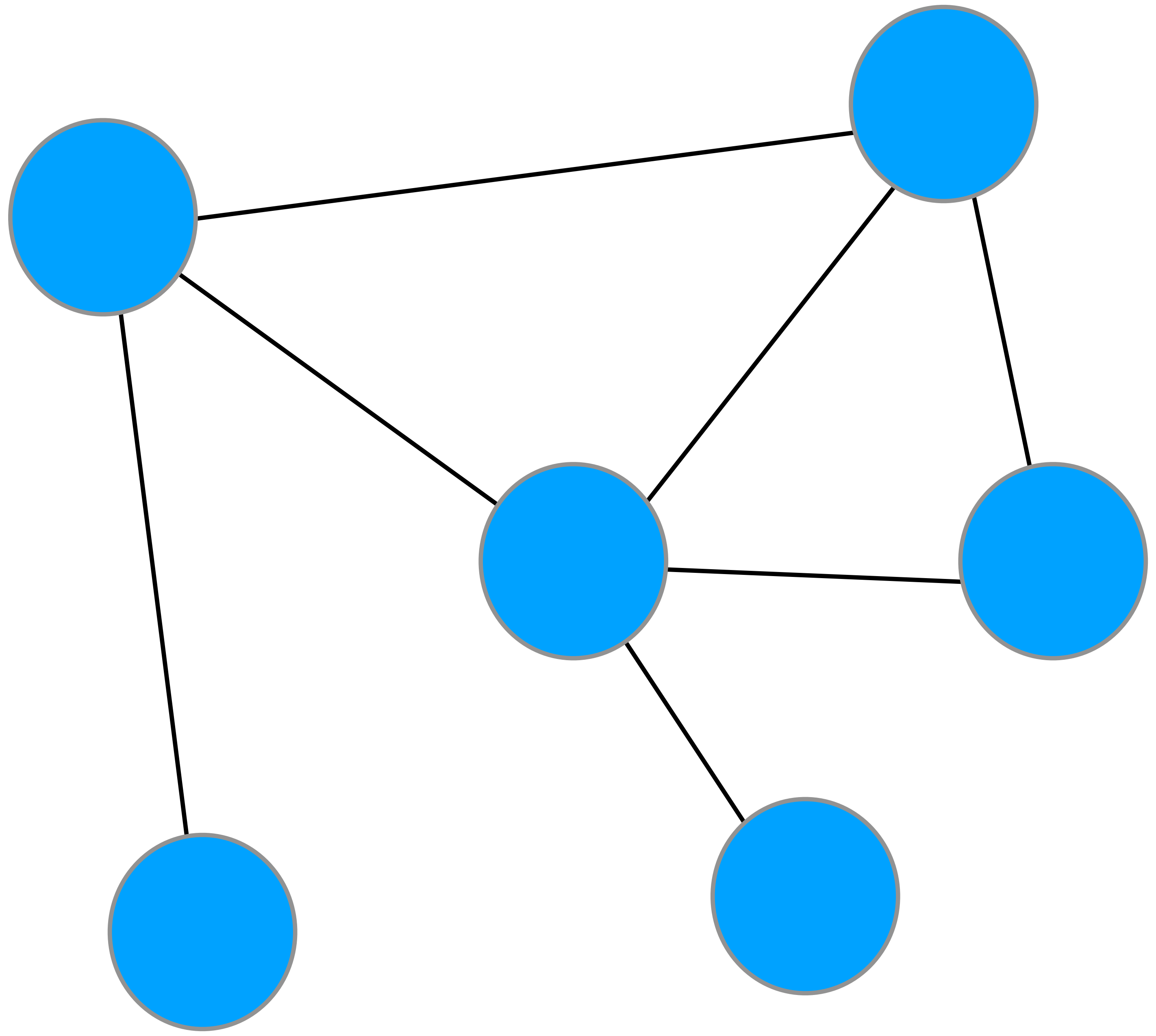


Static Models

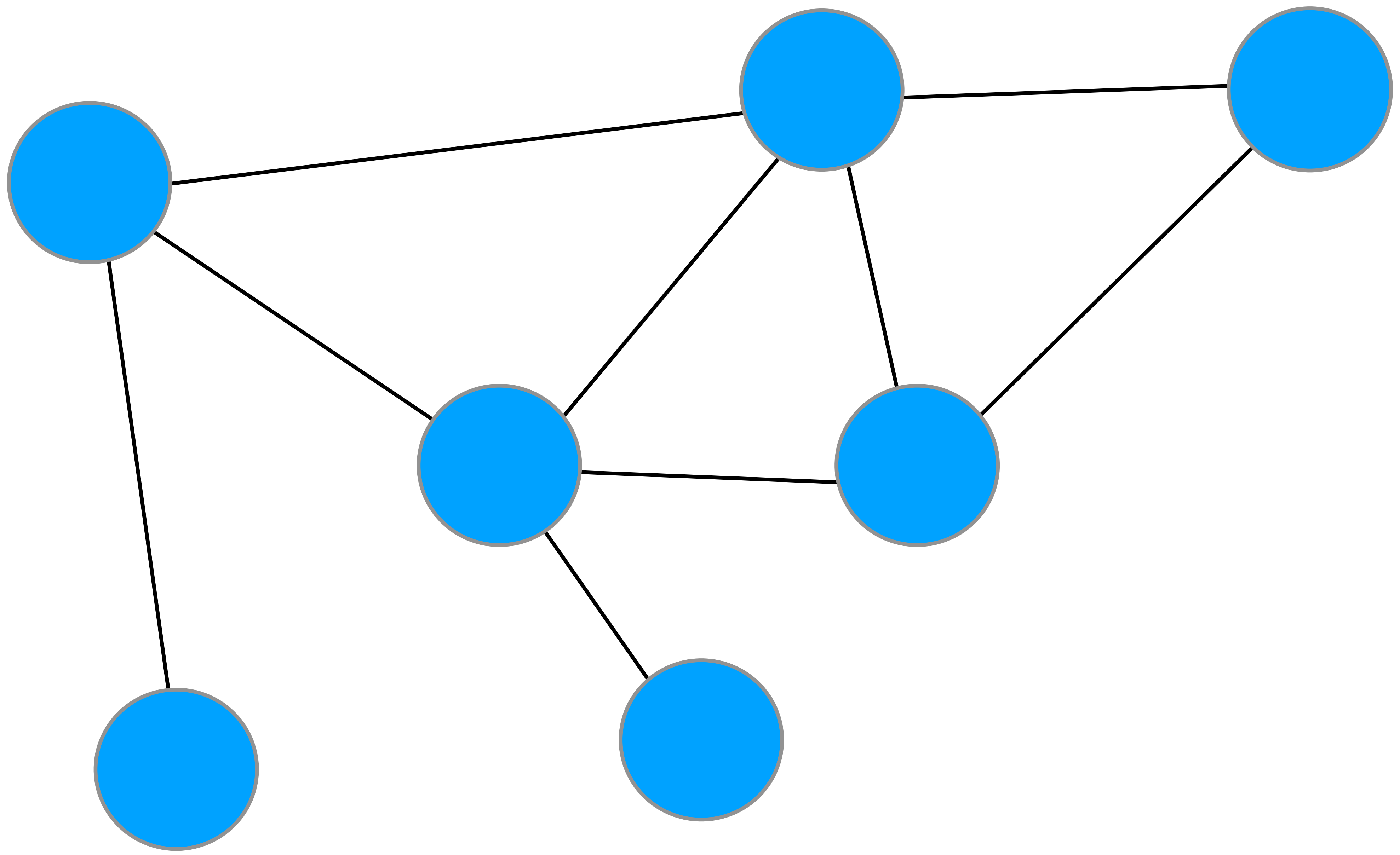


Dynamic Models

Models for network topology

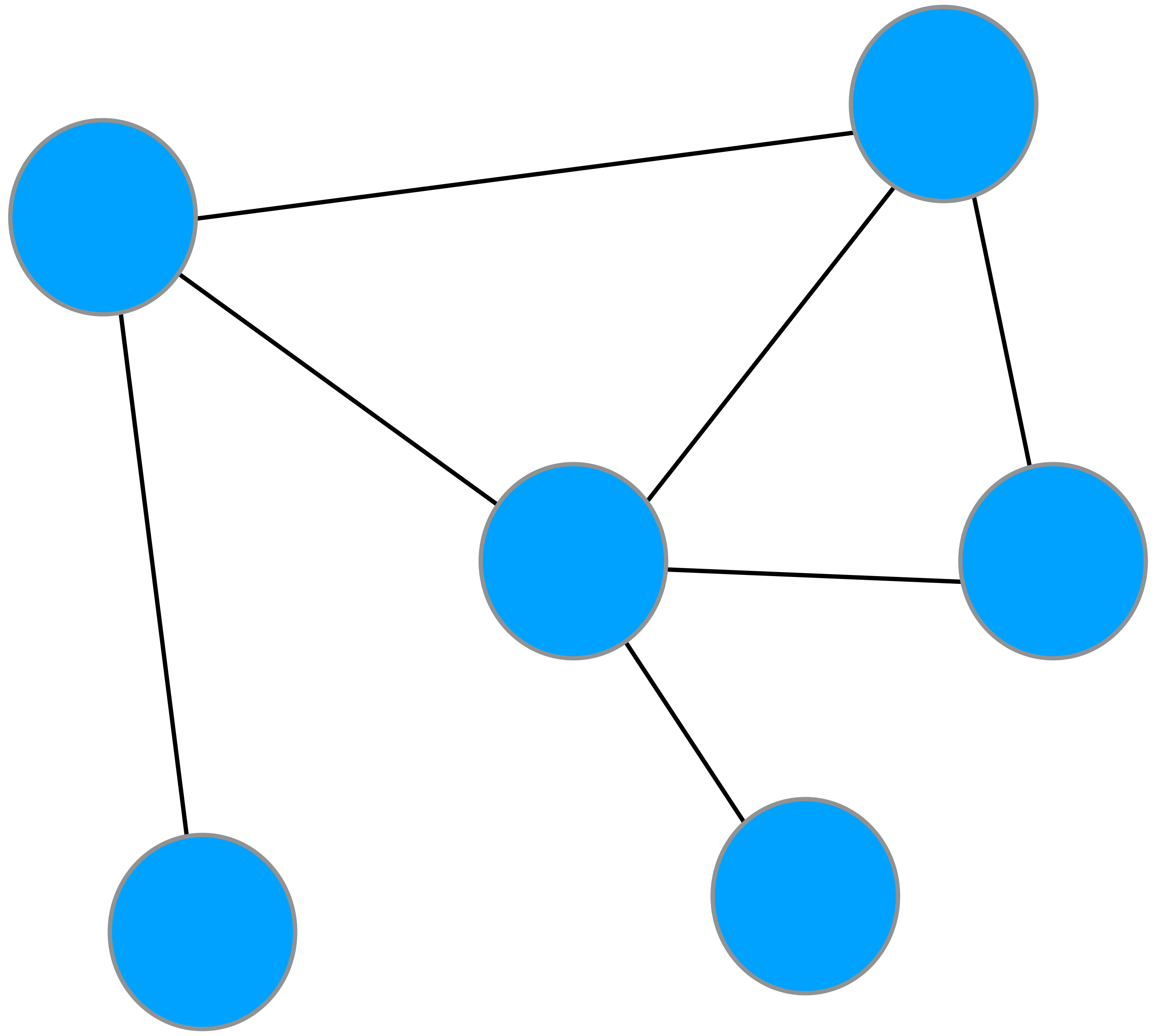


Static Models

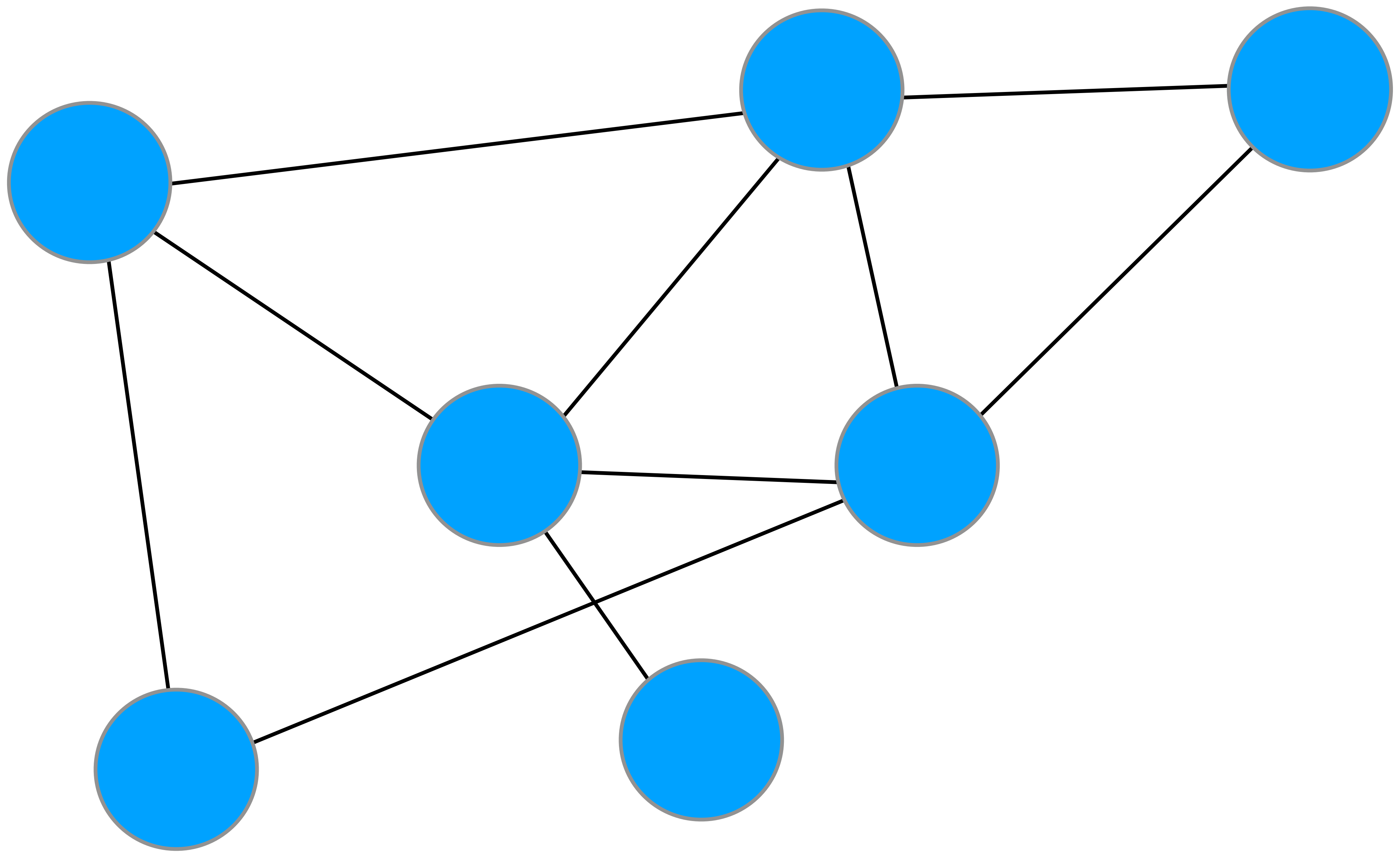


Dynamic Models

Models for network topology

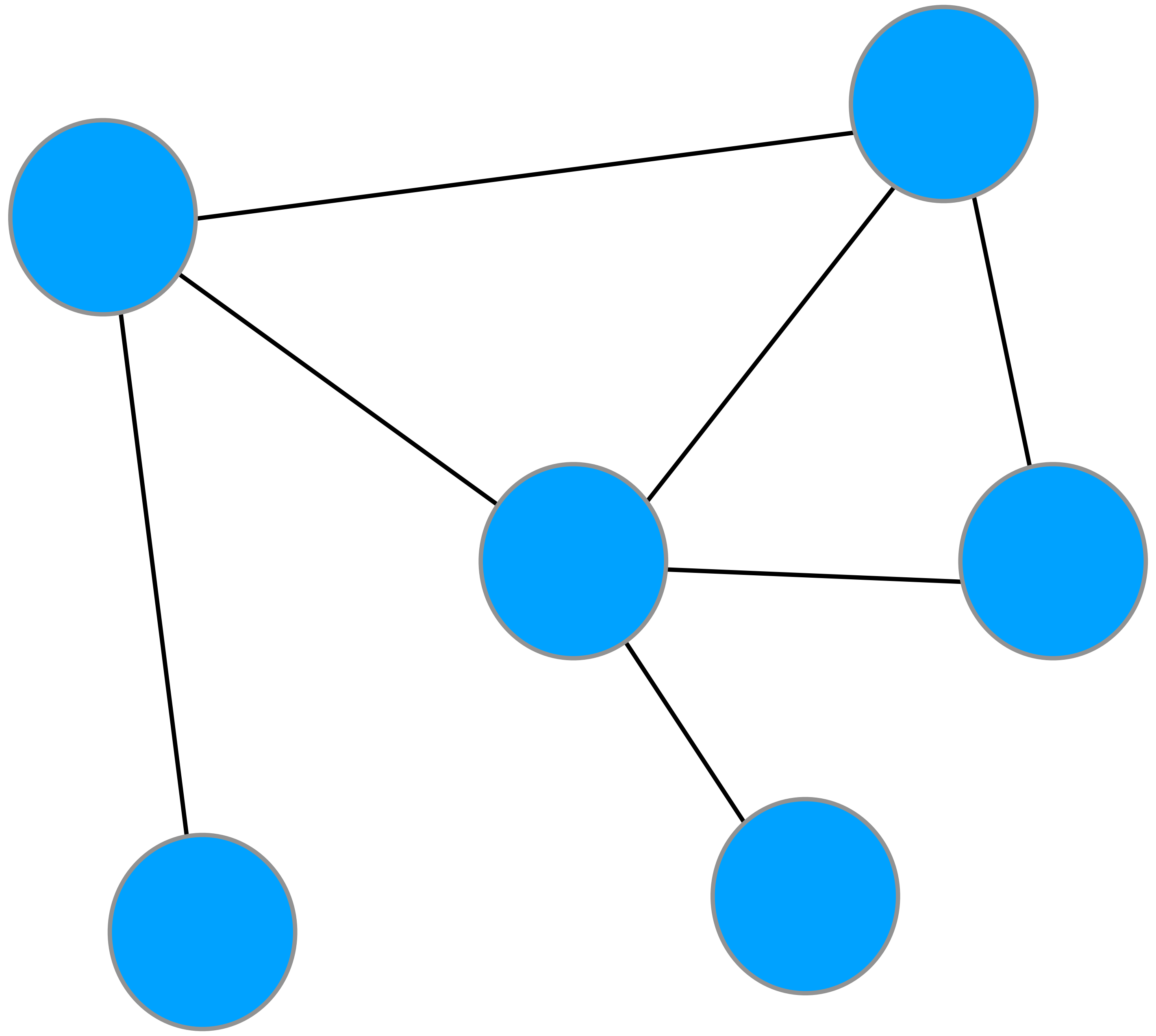


Static Models

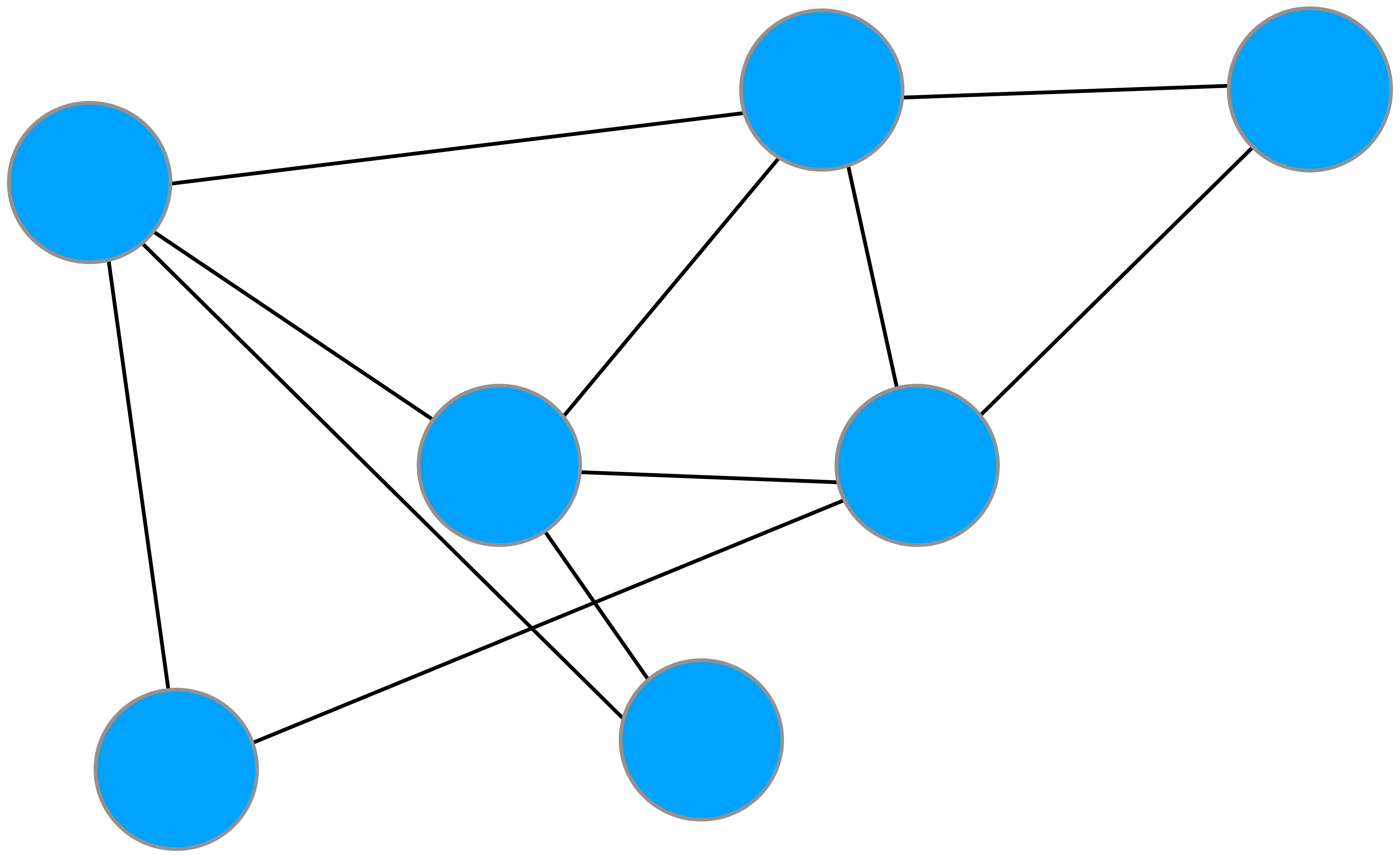


Dynamic Models

Models for network topology

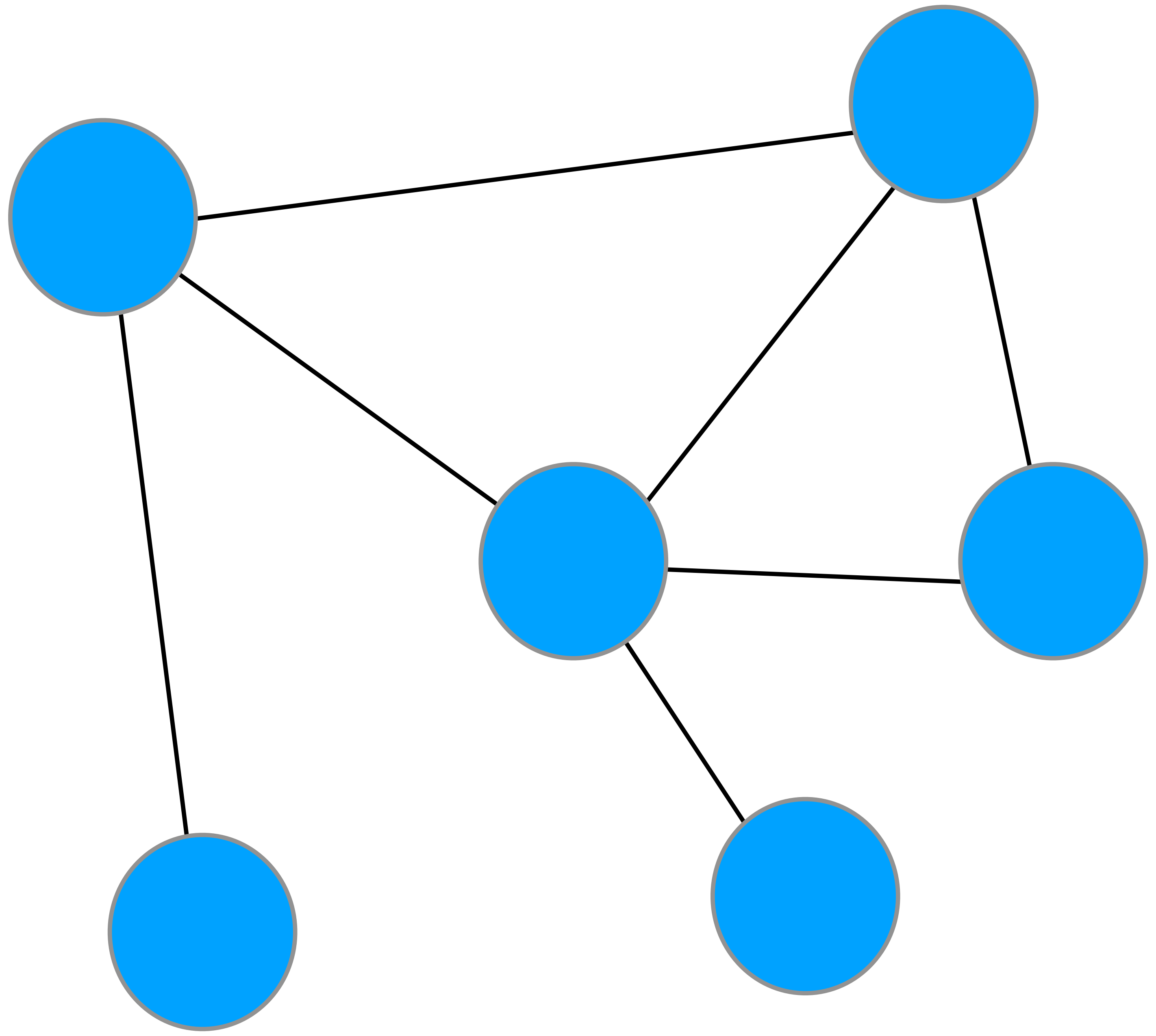


Static Models

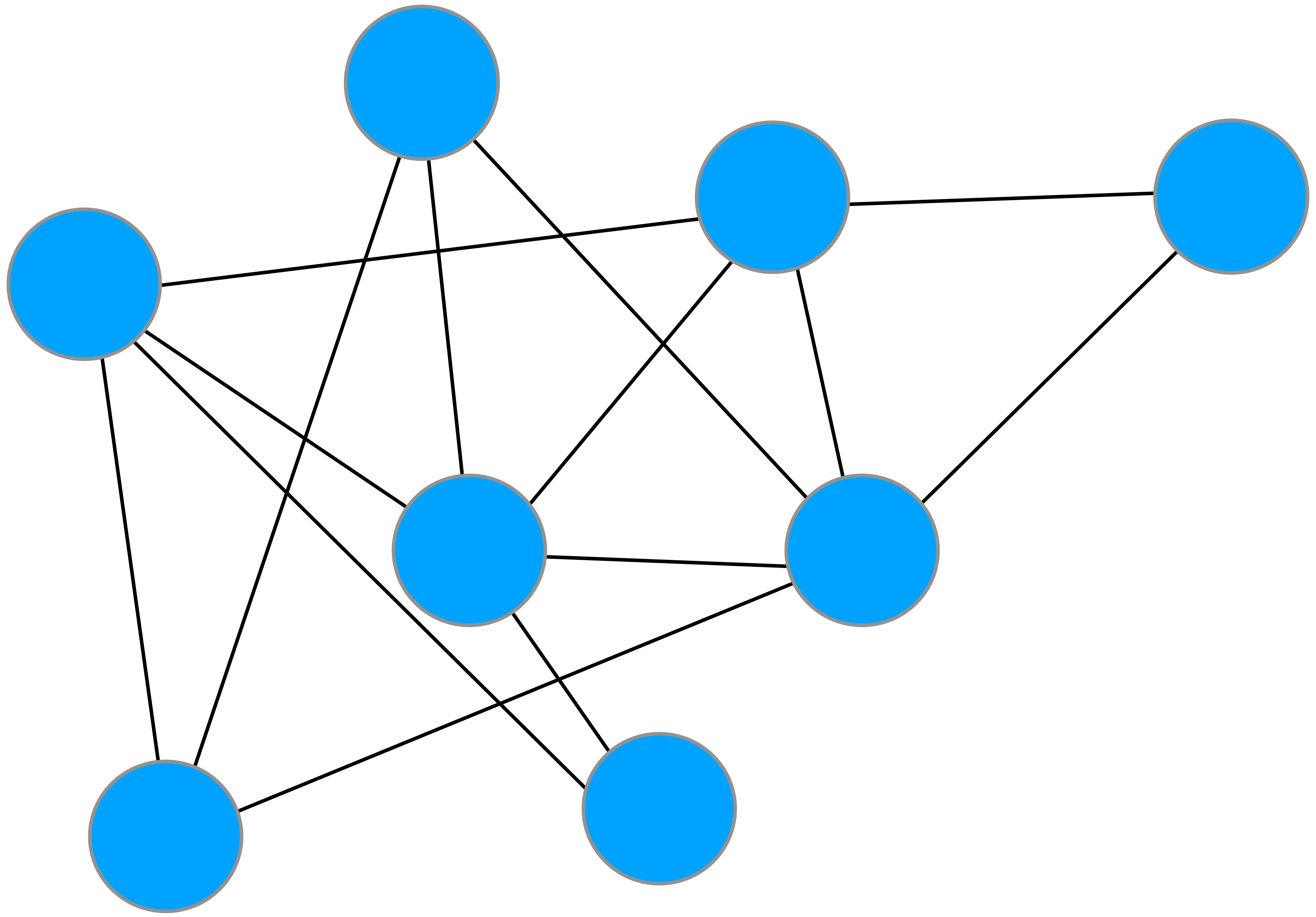


Dynamic Models

Models for network topology

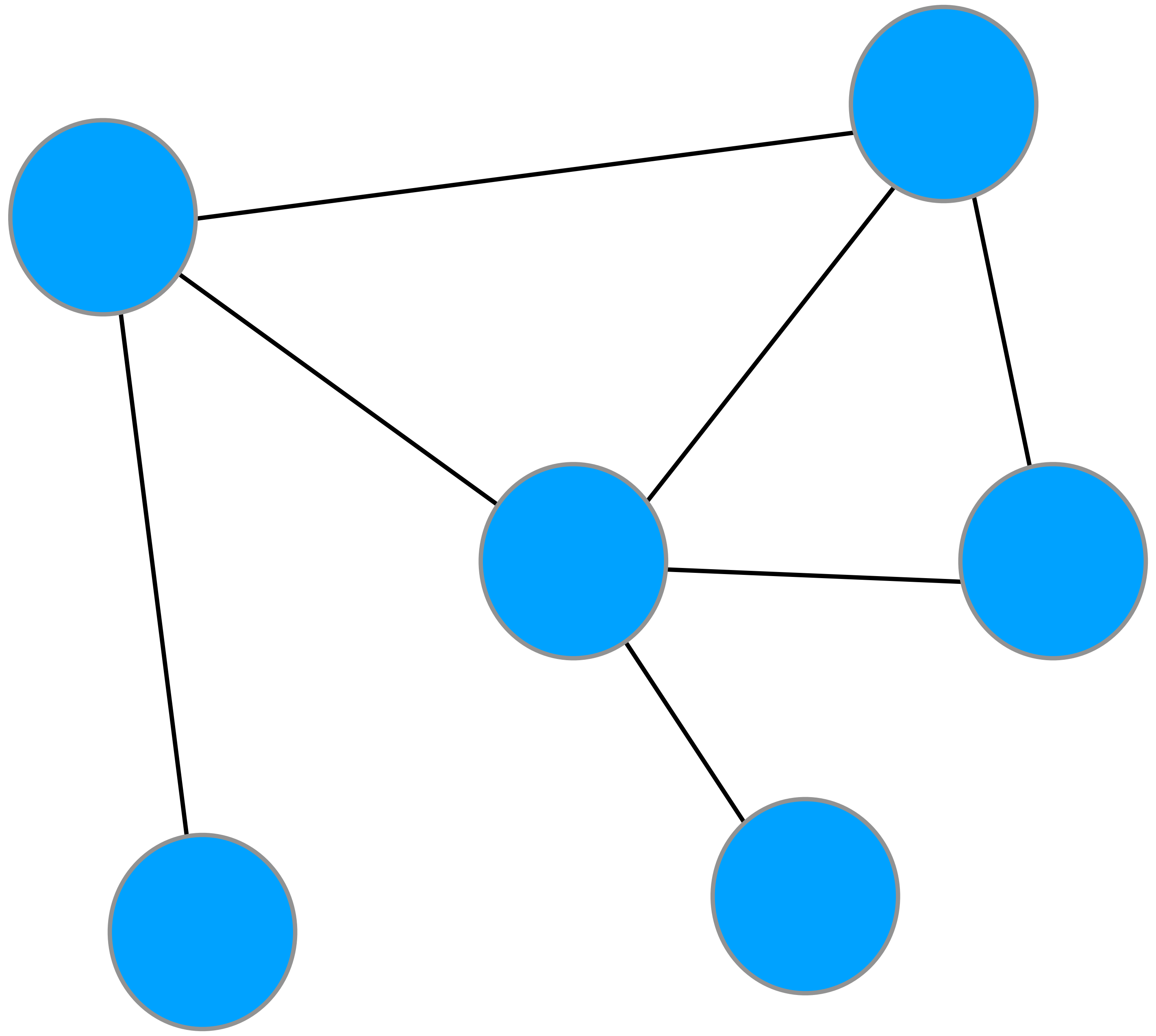


Static Models

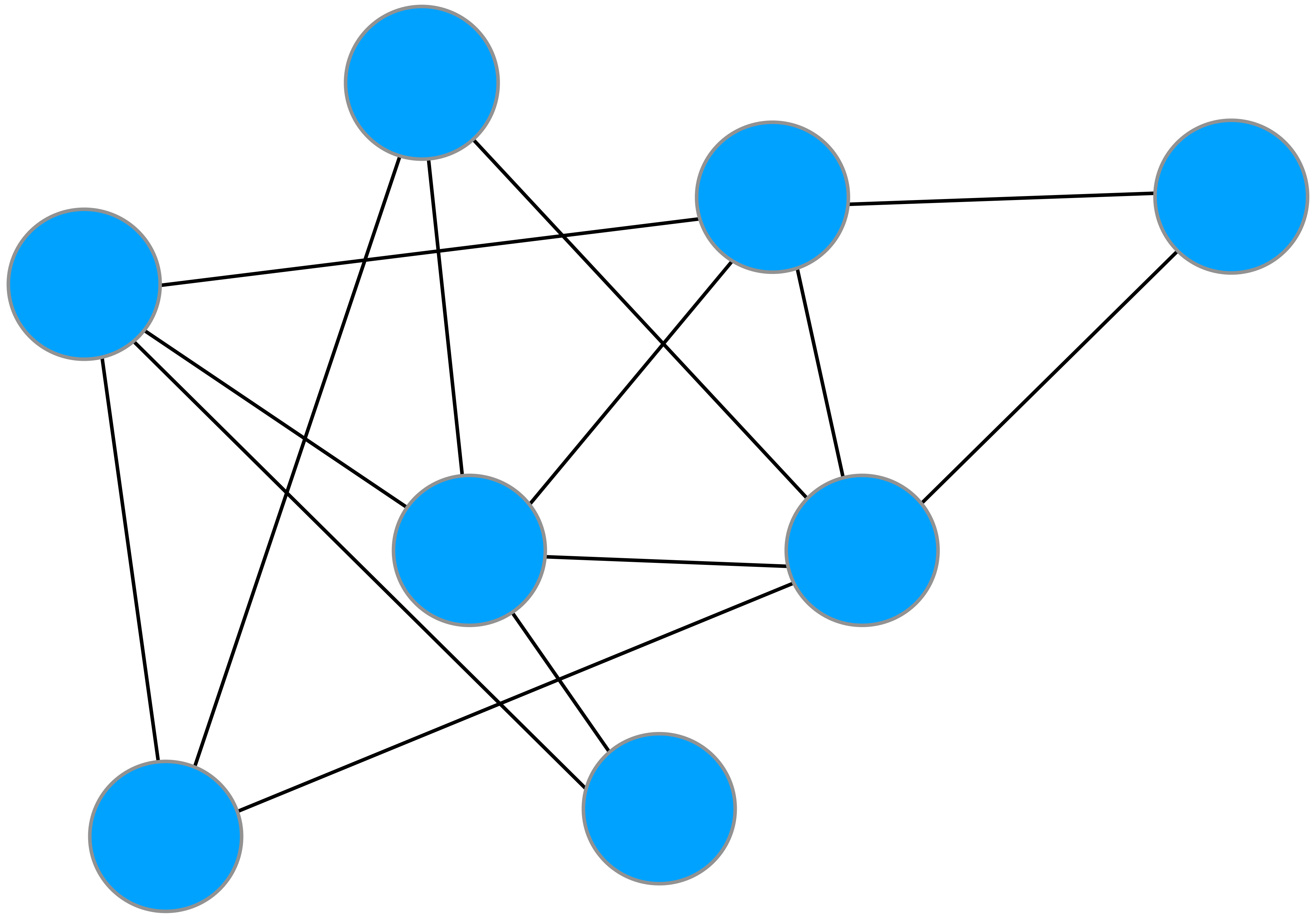


Dynamic Models

Models for network topology



Static Models



Dynamic Models

Why dynamic: Network Changepoints

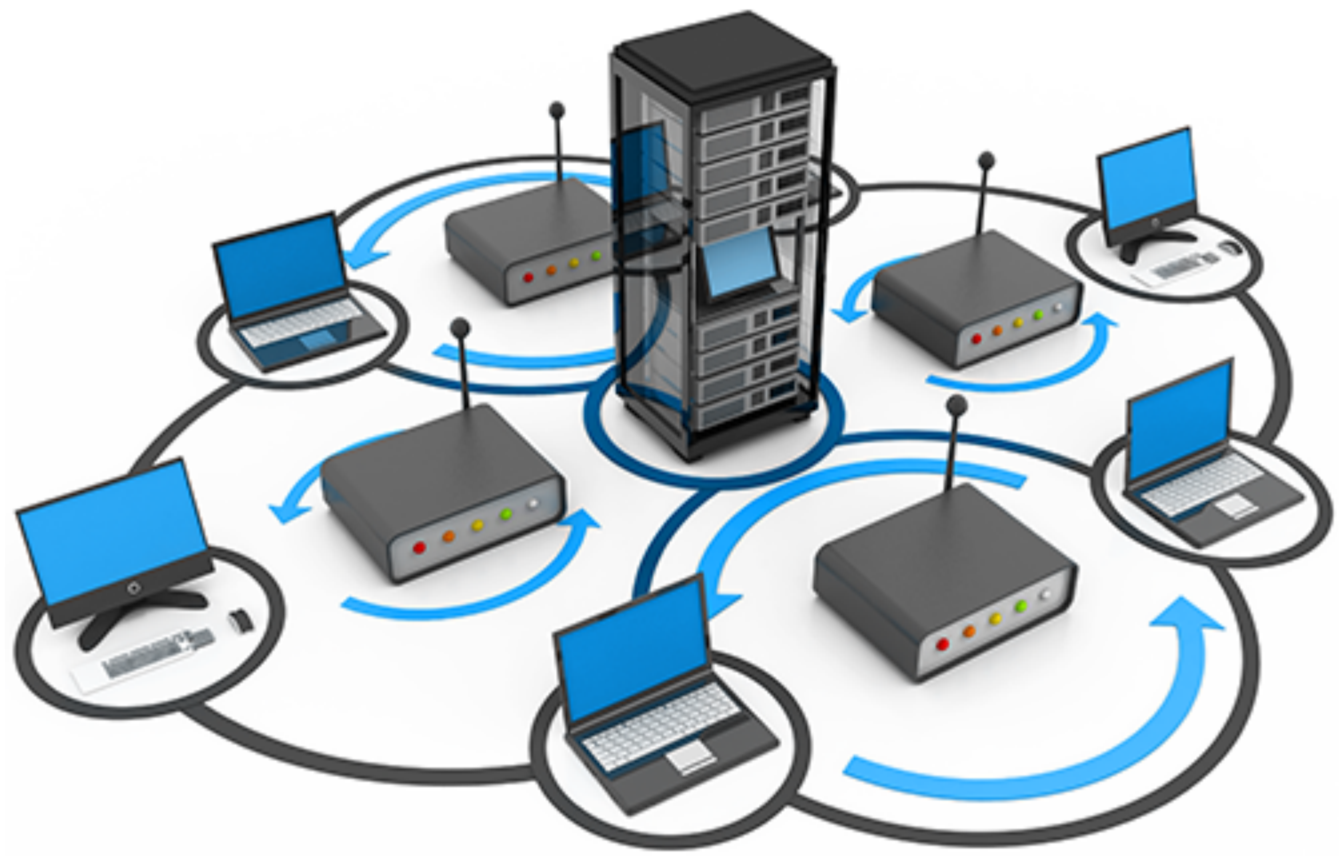
Why dynamic: Network Changepoints

Forks in the blockchain resulting in splitting of transaction network

Why dynamic: Network Changepoints

Forks in the blockchain resulting in splitting of transaction network

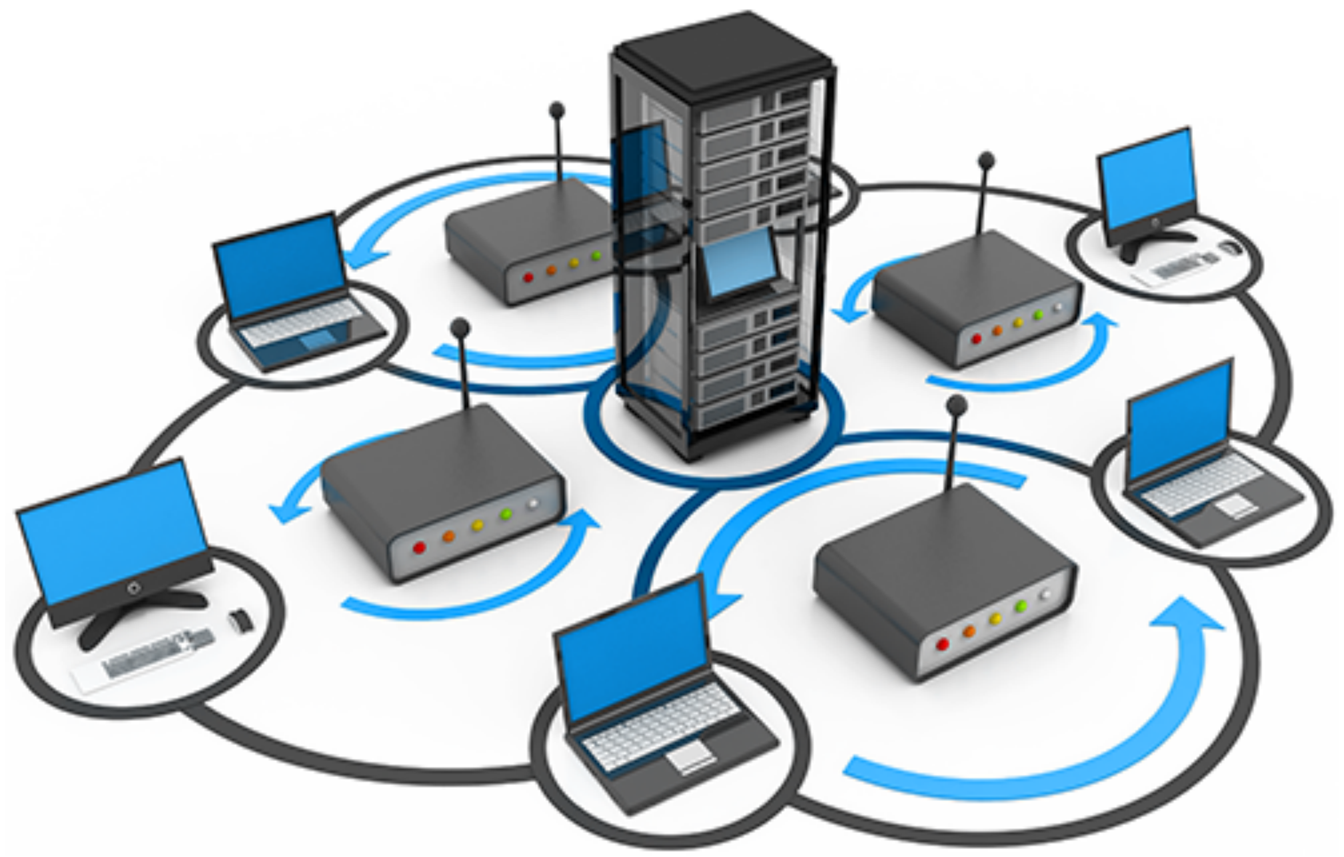
Emergence of IXPs changing peering behaviour



Why dynamic: Network Changepoints

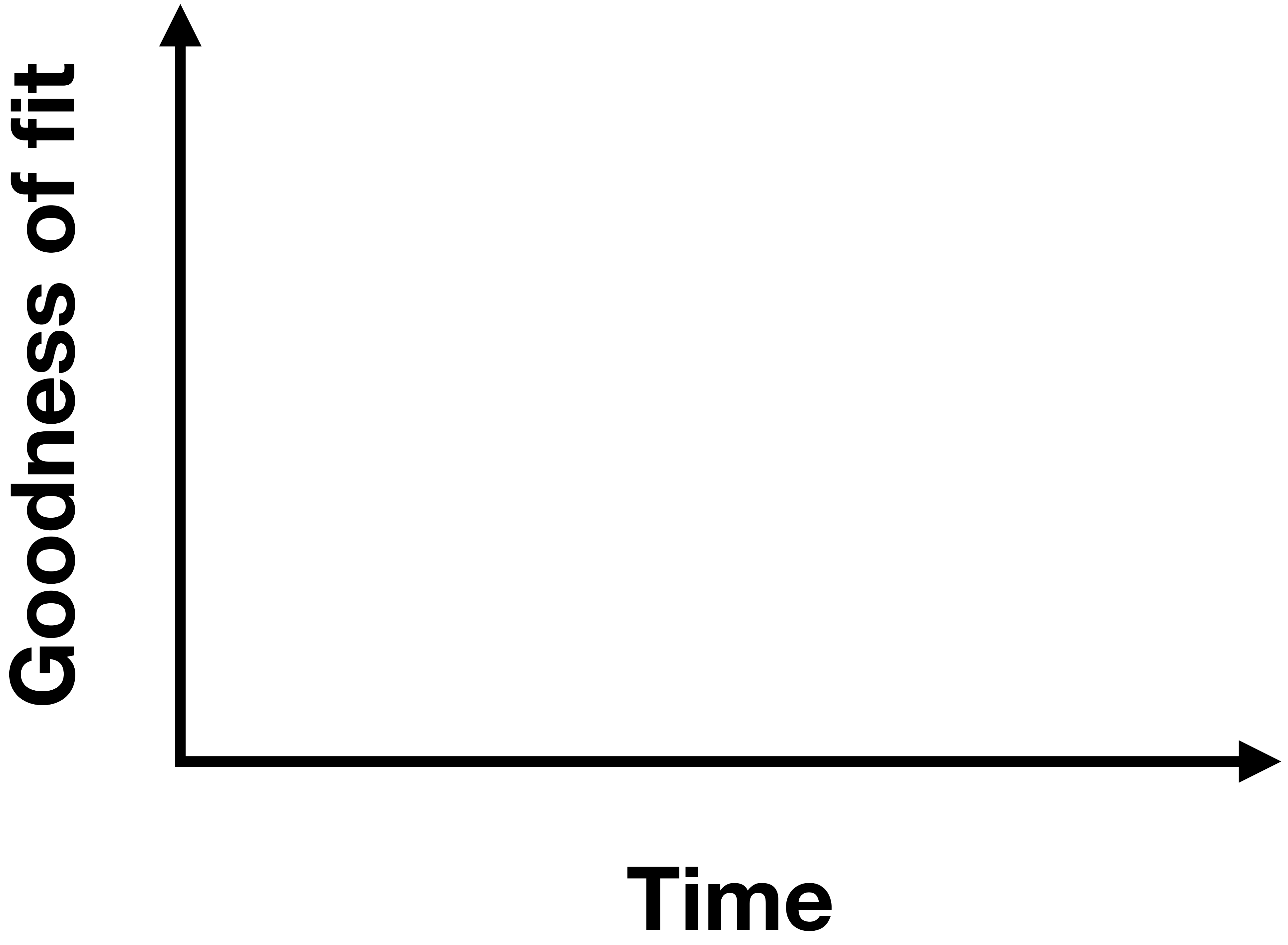
Forks in the blockchain resulting in splitting of transaction network

Emergence of IXPs changing peering behaviour



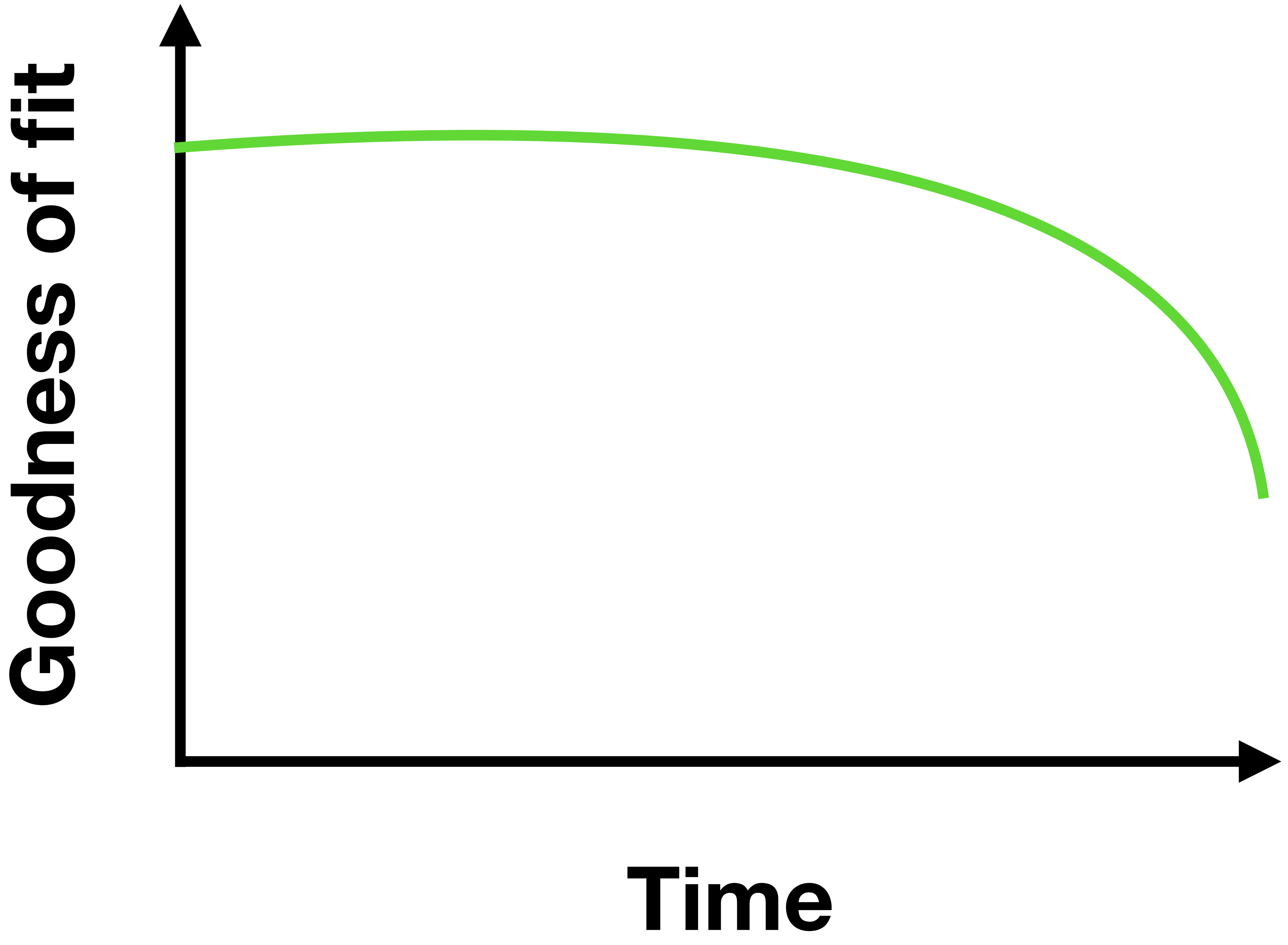
New features which change way in which users connect and interact with each other

Informal idea using network topology models



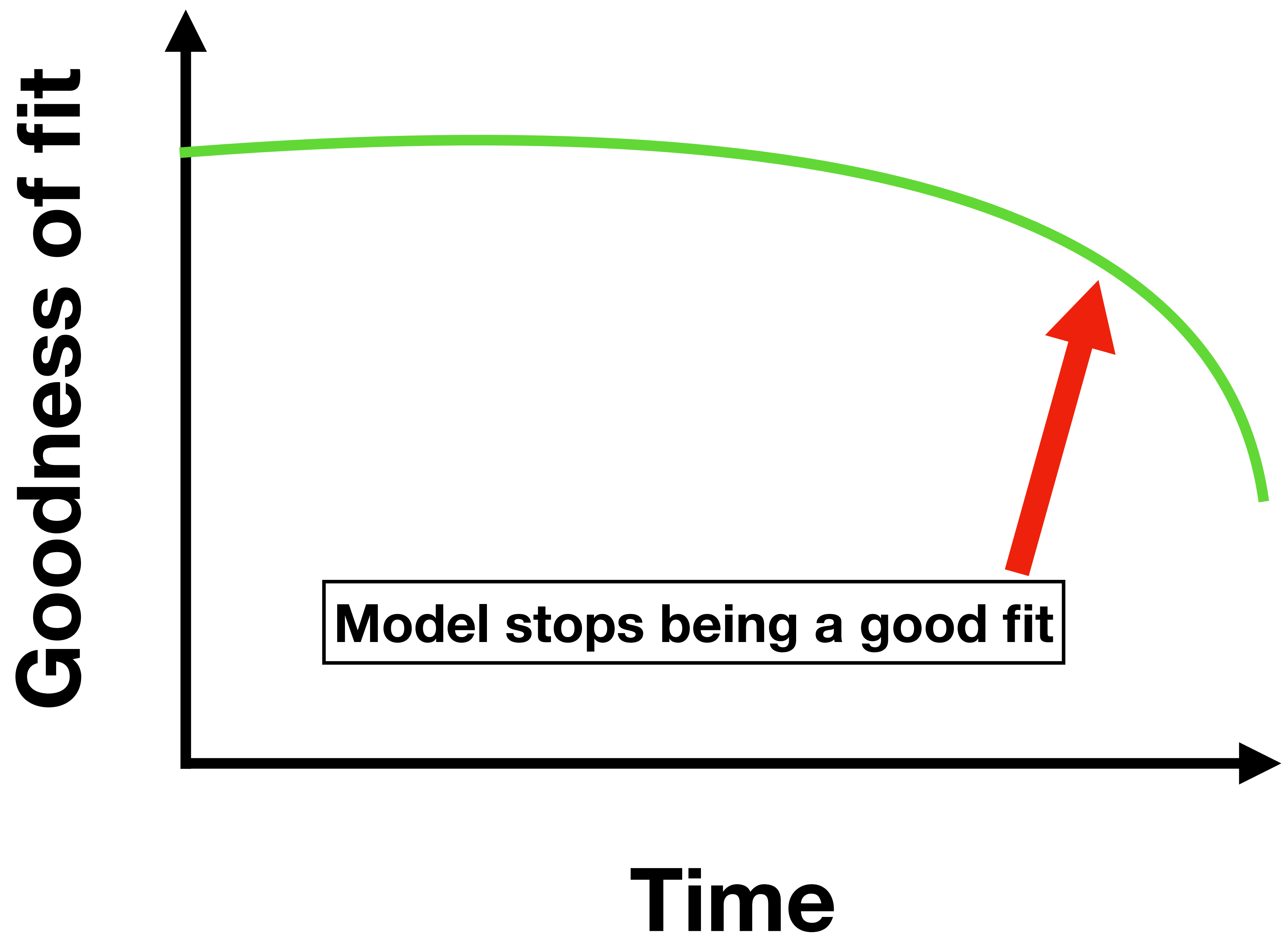
Suppose we have a sound way of monitoring the goodness of fit of a model for a network that is evolving in time

Informal idea using network topology models



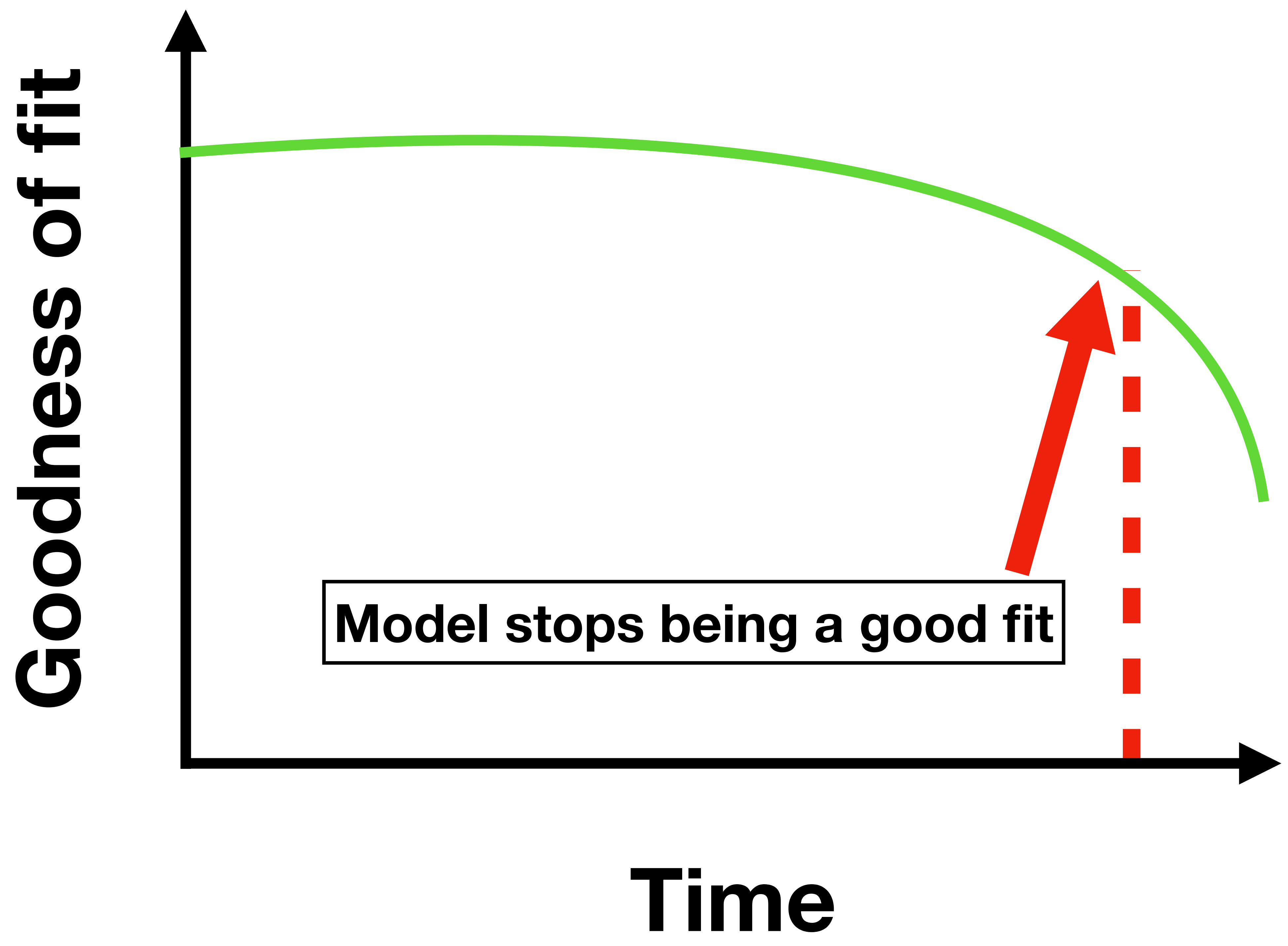
Suppose we have a sound way of monitoring the goodness of fit of a model for a network that is evolving in time

Informal idea using network topology models



Suppose we have a sound way of monitoring the goodness of fit of a model for a network that is evolving in time

Informal idea using network topology models



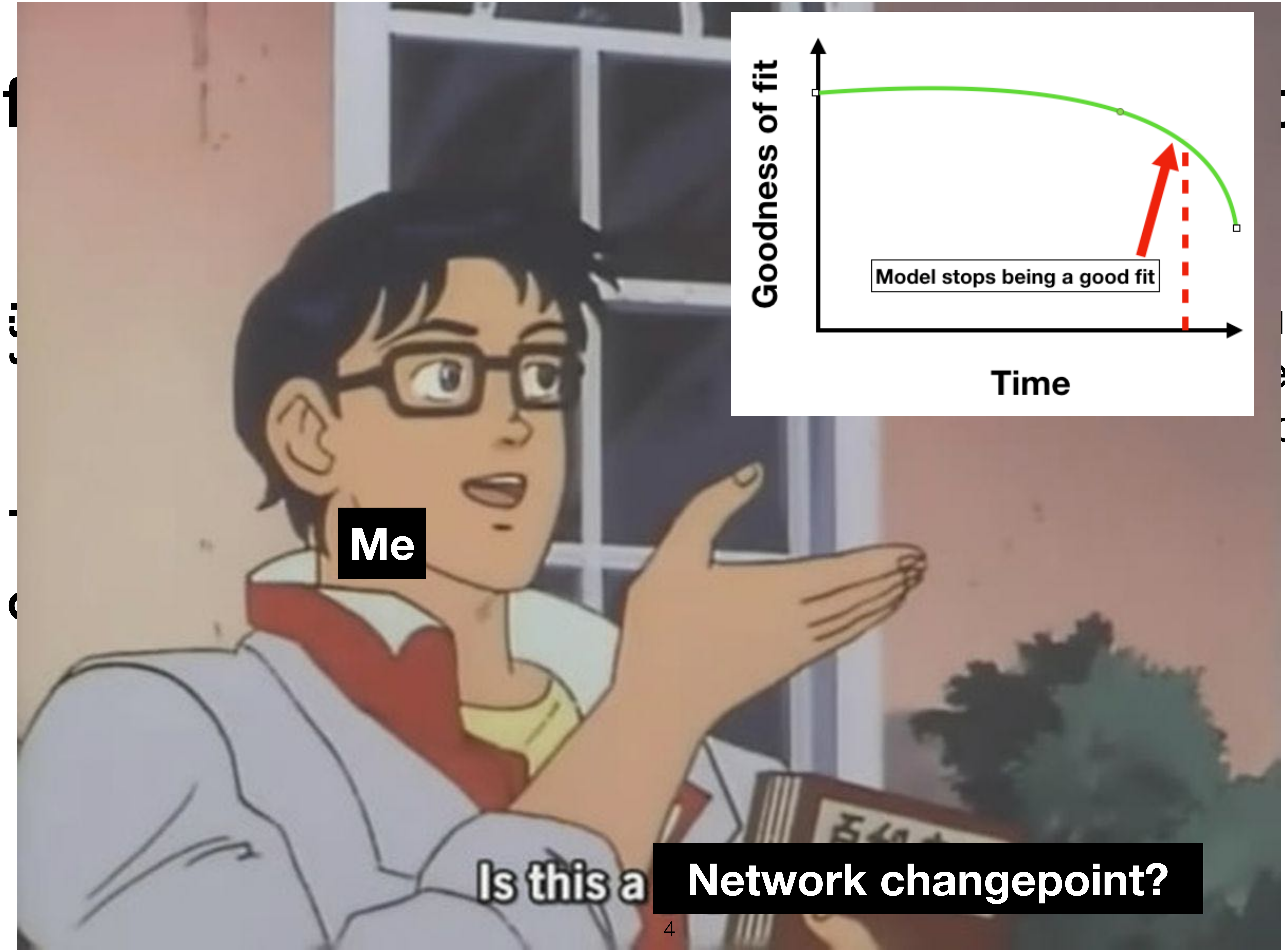
Suppose we have a sound way of monitoring the goodness of fit of a model for a network that is evolving in time

Inf dels

und

e

del



First steps: What is a ‘good model’ for a given evolving network?

First steps: What is a ‘good model’ for a given evolving network?

Problems:

First steps: What is a 'good model' for a given evolving network?

Problems:

Many statistics you can use, how do you decide which are the important ones?

First steps: What is a 'good model' for a given evolving network?

Problems:

Many statistics you can use, how do you decide which are the important ones?

Certain statistics highly variable from one realisation to another e.g. maximum degree

First steps: What is a ‘good model’ for a given evolving network?

Problems:

Many statistics you can use, how do you decide which are the important ones?

Generating networks takes time, especially when trying different parameters

Certain statistics highly variable from one realisation to another e.g. maximum degree

First steps: What is a 'good model' for a given evolving network?

Problems:

Many statistics you can use, how do you decide which are the important ones?

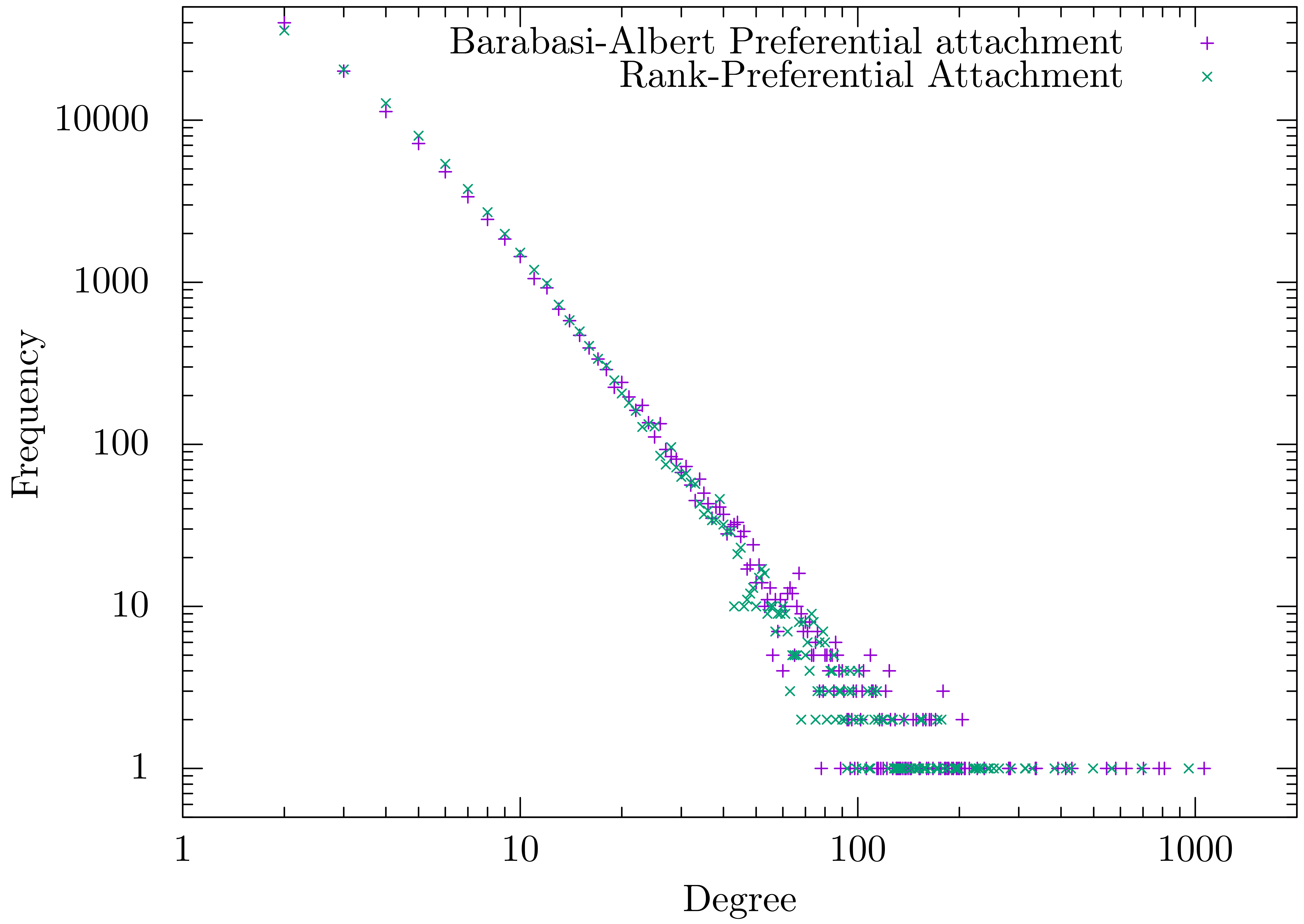
Generating networks takes time, especially when trying different parameters

Certain statistics highly variable from one realisation to another e.g. maximum degree

Different explanatory models might generate networks with similar/same statistics

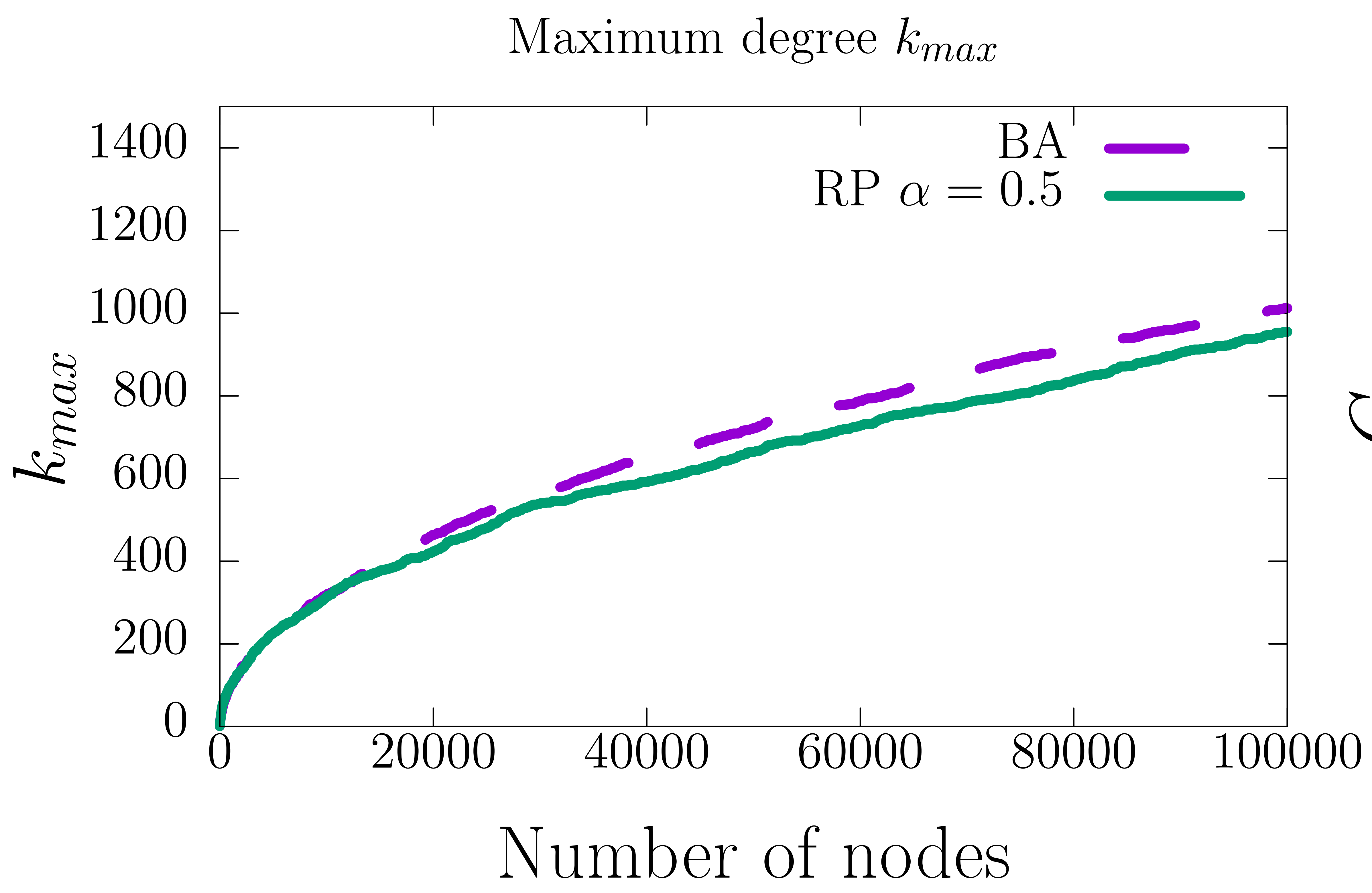
Example: Barabasi-Albert vs Rank-Preference

Example: Barabasi-Albert vs Rank-Preference

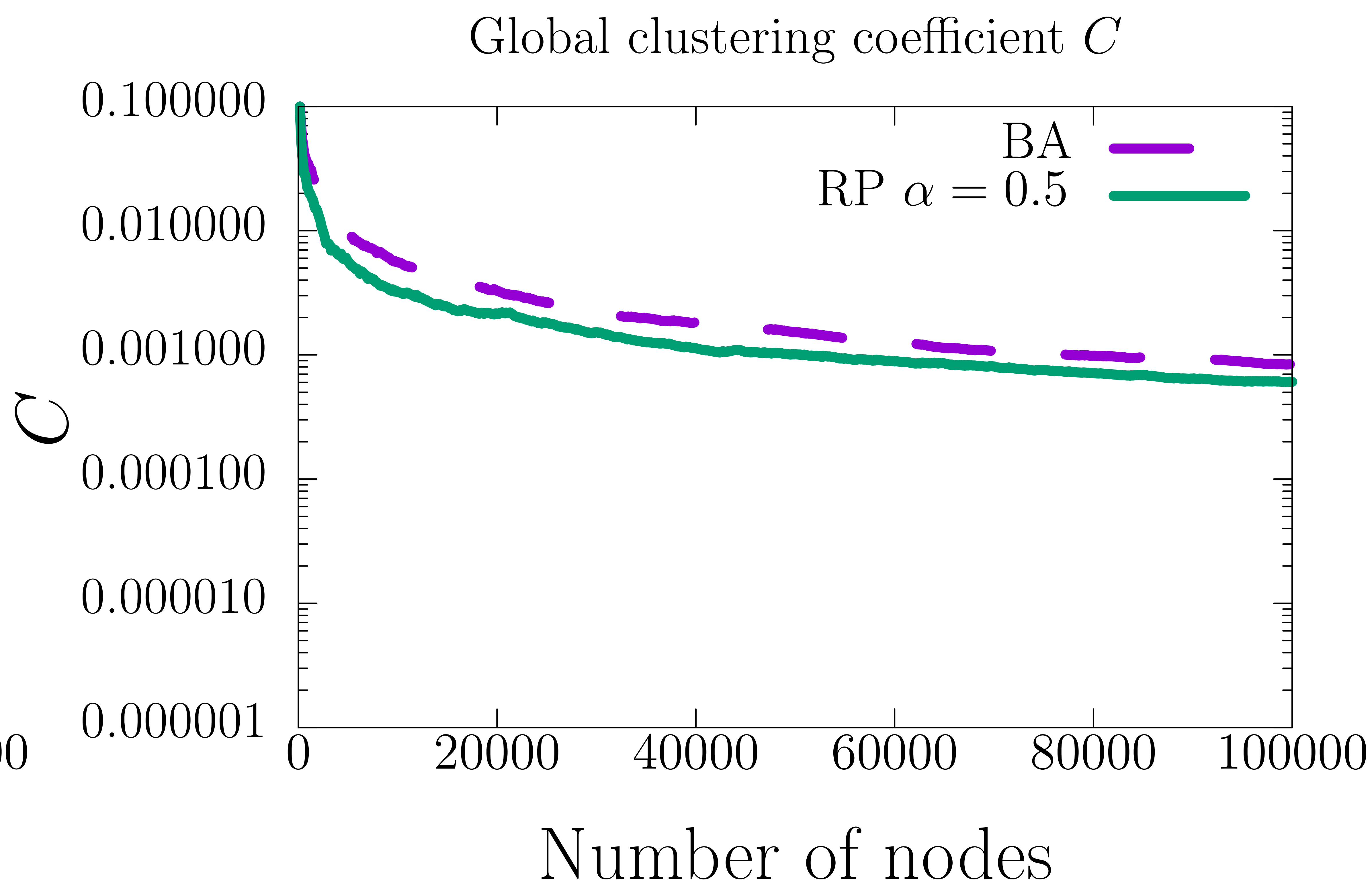


Degree distribution of generated networks

Example: Barabasi-Albert vs Rank-Preference

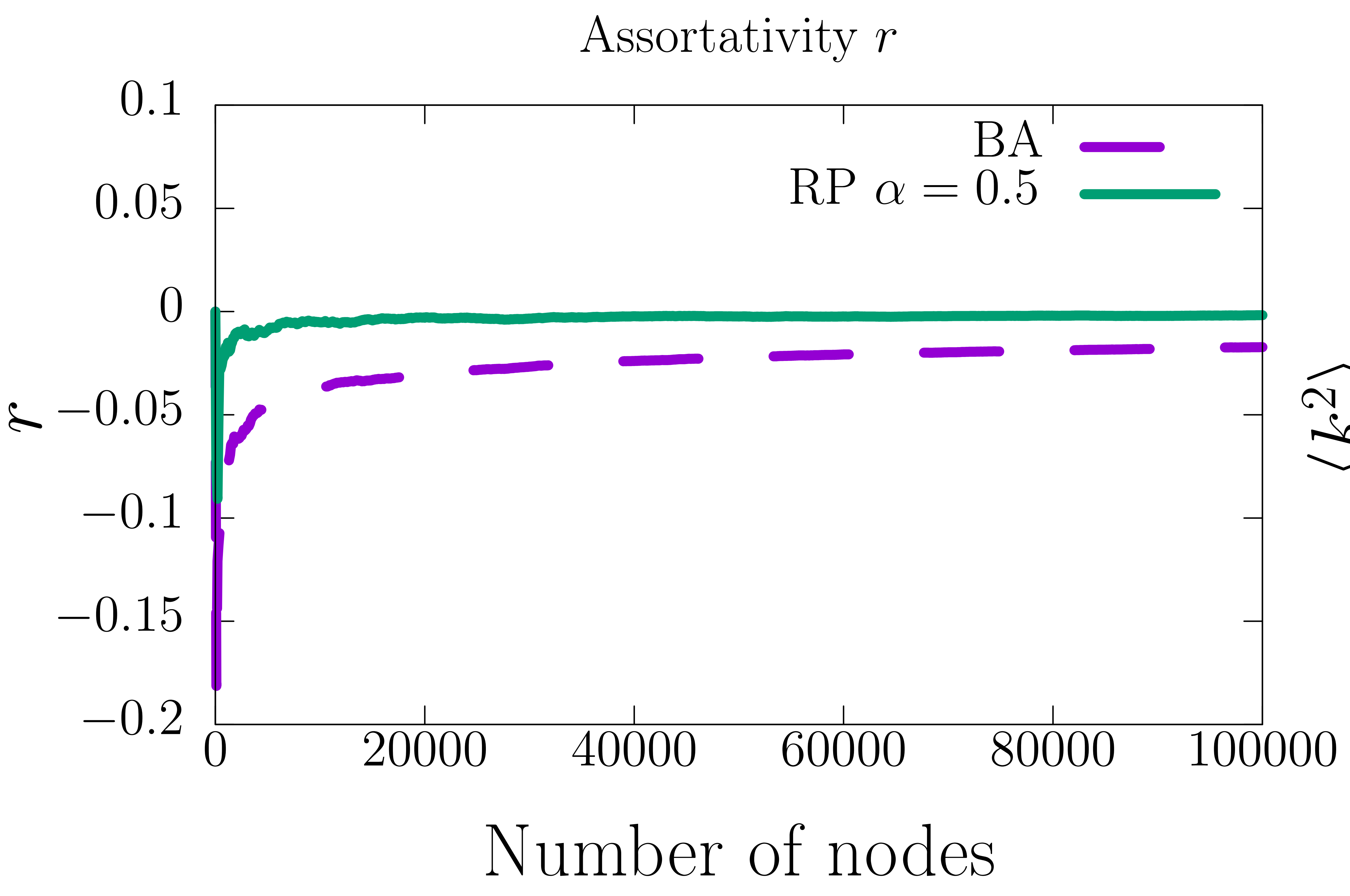


Maximum Degree

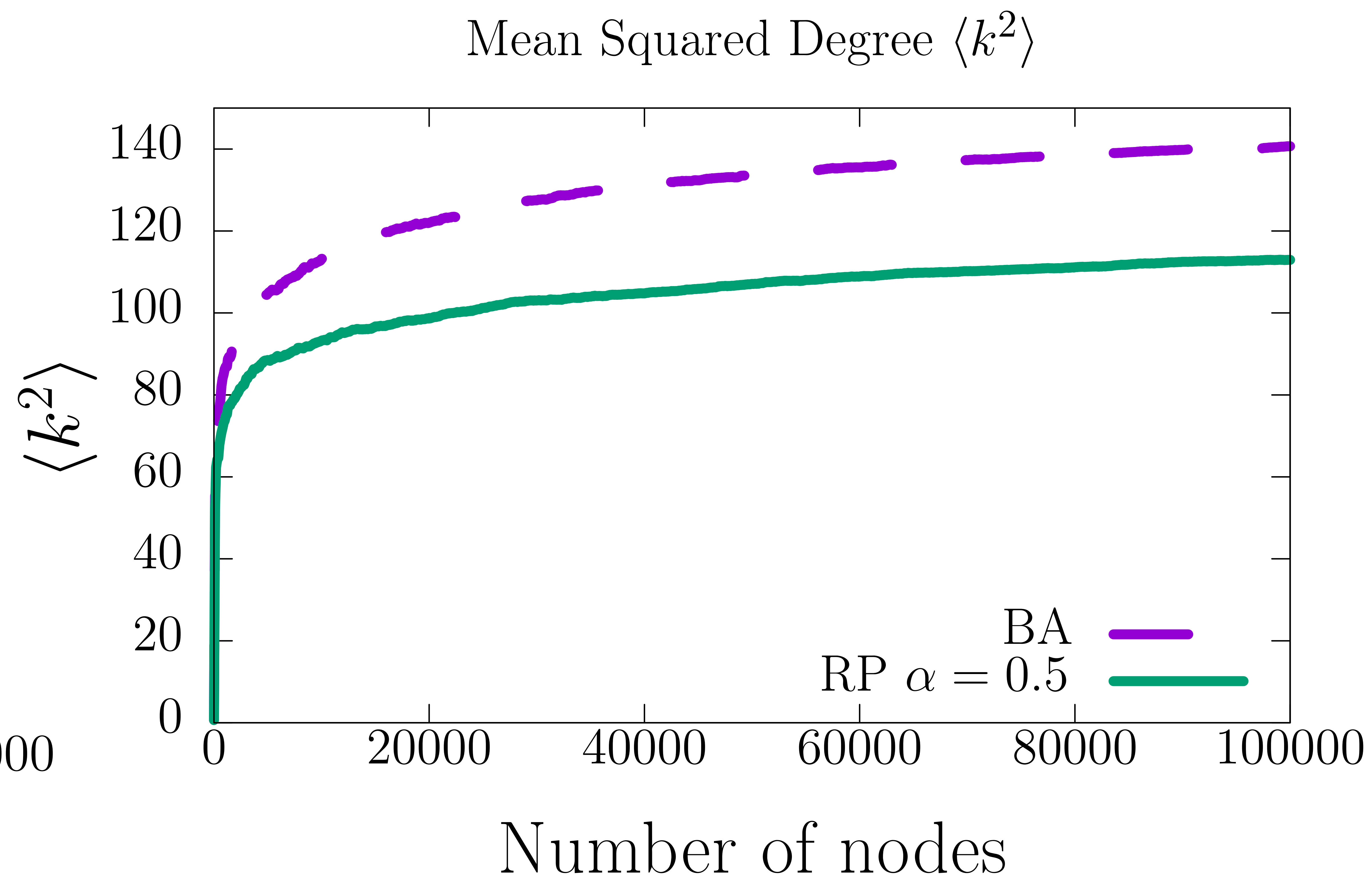


Global Clustering Coefficient

Example: Barabasi-Albert vs Rank-Preference



Degree Assortativity



Mean Squared Degree

Better Measure: Model Likelihood

[R. G. Clegg, B. Parker, M. Rio
Likelihood based assessment of dynamic networks (2016)]

Better Measure: Model Likelihood

Evolving network $G(t)$

Candidate probabilistic model $M(\theta)$

[R. G. Clegg, B. Parker, M. Rio *Likelihood based assessment of dynamic networks* (2016)]

Better Measure: Model Likelihood

Evolving network $G(t)$

Candidate probabilistic model $M(\theta)$

[R. G. Clegg, B. Parker, M. Rio *Likelihood based assessment of dynamic networks* (2016)]

Likelihood of model: Probability of generating entire evolution of G from model $M(\theta)$

Better Measure: Model Likelihood

Many statistics you can use, how do you decide which are the important ones?

Generating networks takes time, especially when trying different parameters

Certain statistics highly variable from one realisation to another e.g. maximum degree

Different explanatory models might generate networks with similar/same statistics

Better Measure: Model Likelihood

Many statistics you can use, how do you decide which are the

Not relying on single statistics but instead a network-wide measure

variable from one realisation to another e.g. maximum degree

Generating networks takes time, especially when trying different parameters

Different explanatory models might generate networks with similar/same statistics

Better Measure: Model Likelihood

Many statistics you can use, how do you decide which are the

Not relying on single statistics but instead a network-wide measure

variable from one realisation to another e.g. maximum degree

Generating networks takes

Calculating likelihood is much quicker than generating networks and taking measurements

Different explanatory models might generate networks with similar/same statistics

Better Measure: Model Likelihood

Many statistics you can use, how do you decide which are the

Not relying on single statistics but instead a network-wide measure

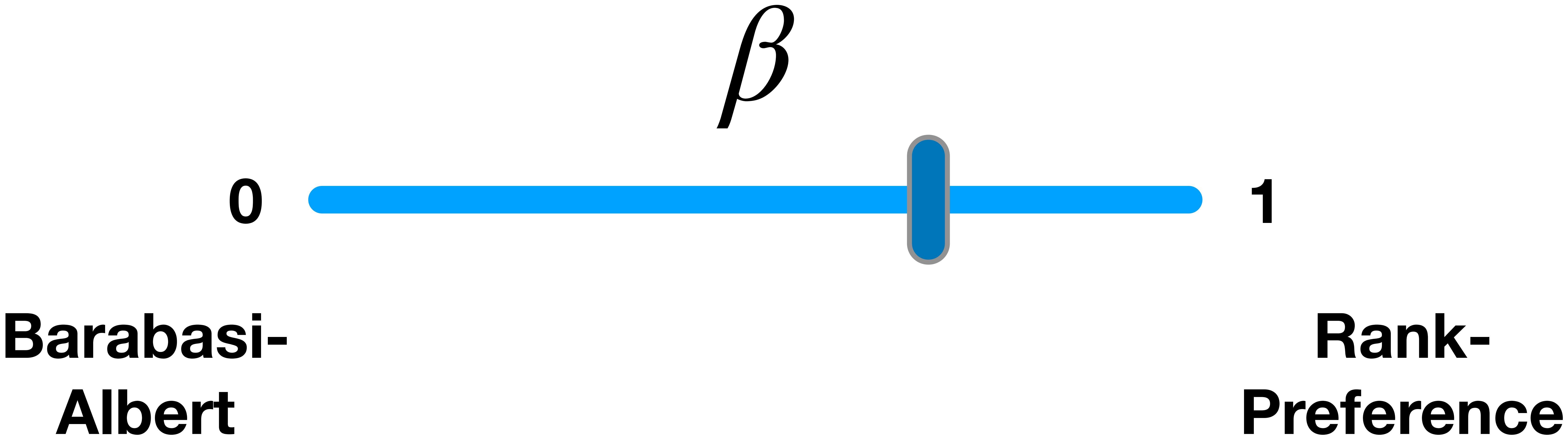
variable from one realisation to another e.g. maximum degree

Generating networks takes

Calculating likelihood is much quicker than generating networks and taking measurements

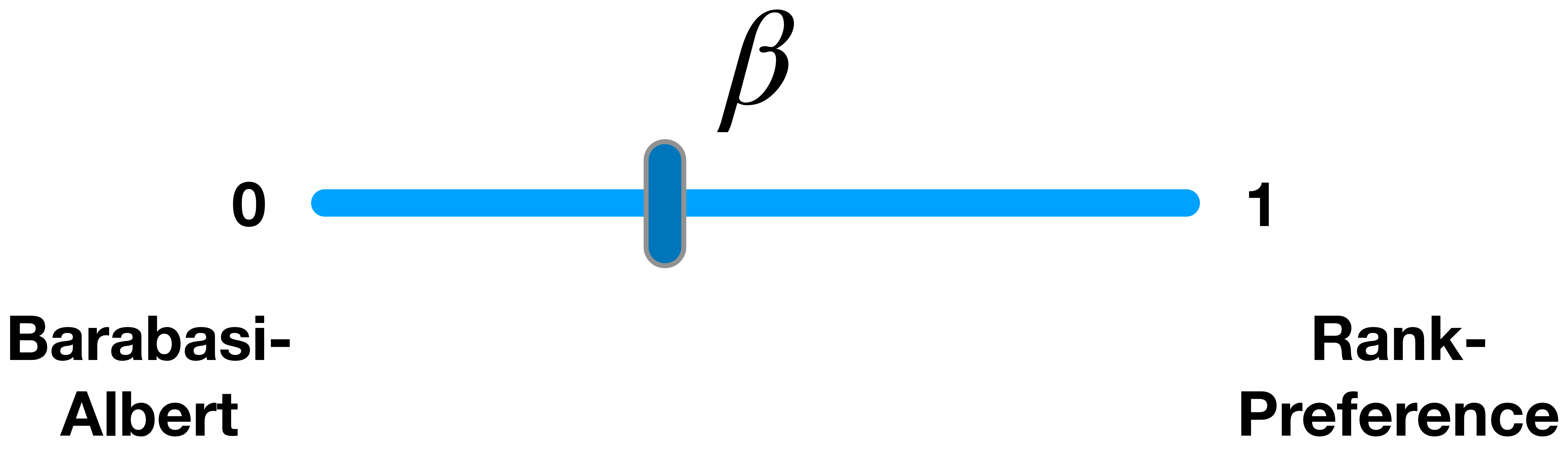
Actually, we can distinguish these using likelihood where single statistics might fail!

Distinguishing networks generated by BA vs RP



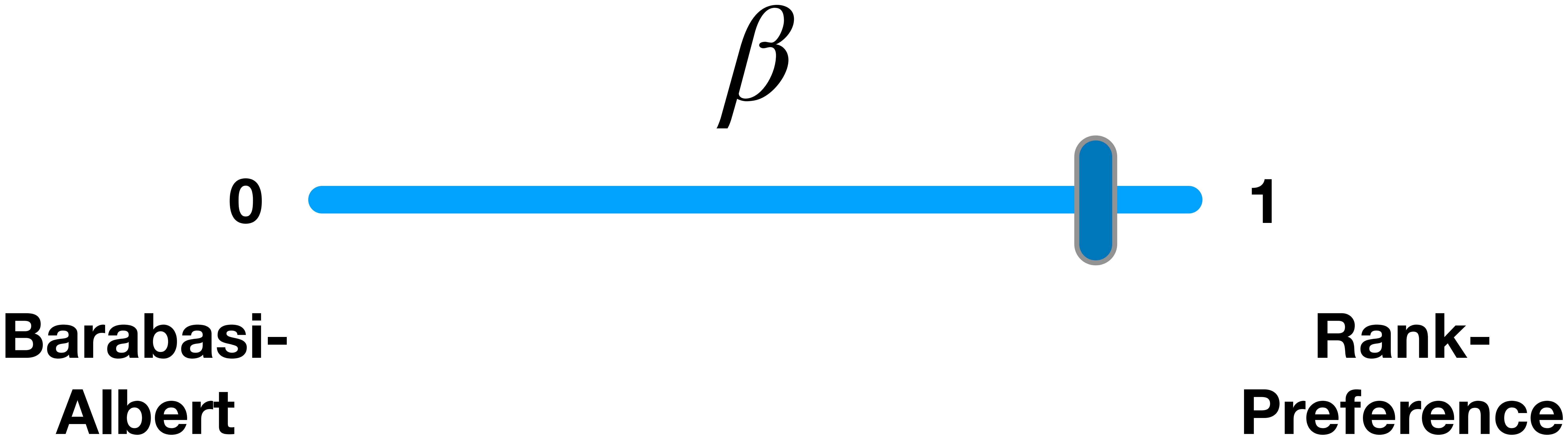
Test: given an artificial network grown that is generated from part BA and part RP, can we recover the proportion each model contributed using model likelihood?

Distinguishing networks generated by BA vs RP



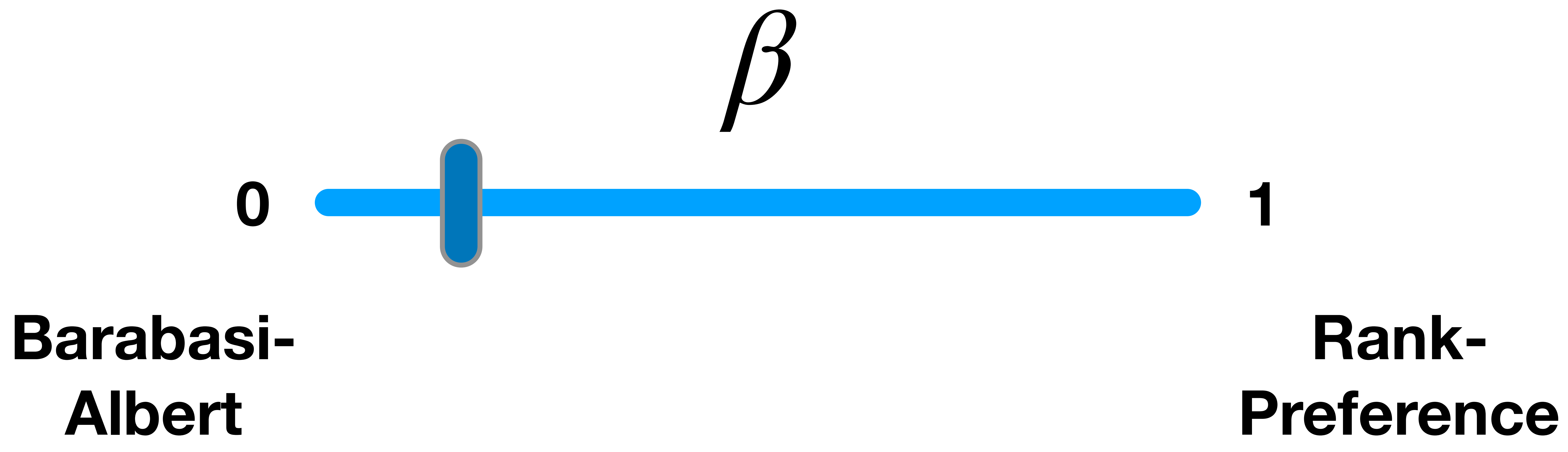
Test: given an artificial network grown that is generated from part BA and part RP, can we recover the proportion each model contributed using model likelihood?

Distinguishing networks generated by BA vs RP



Test: given an artificial network grown that is generated from part BA and part RP, can we recover the proportion each model contributed using model likelihood?

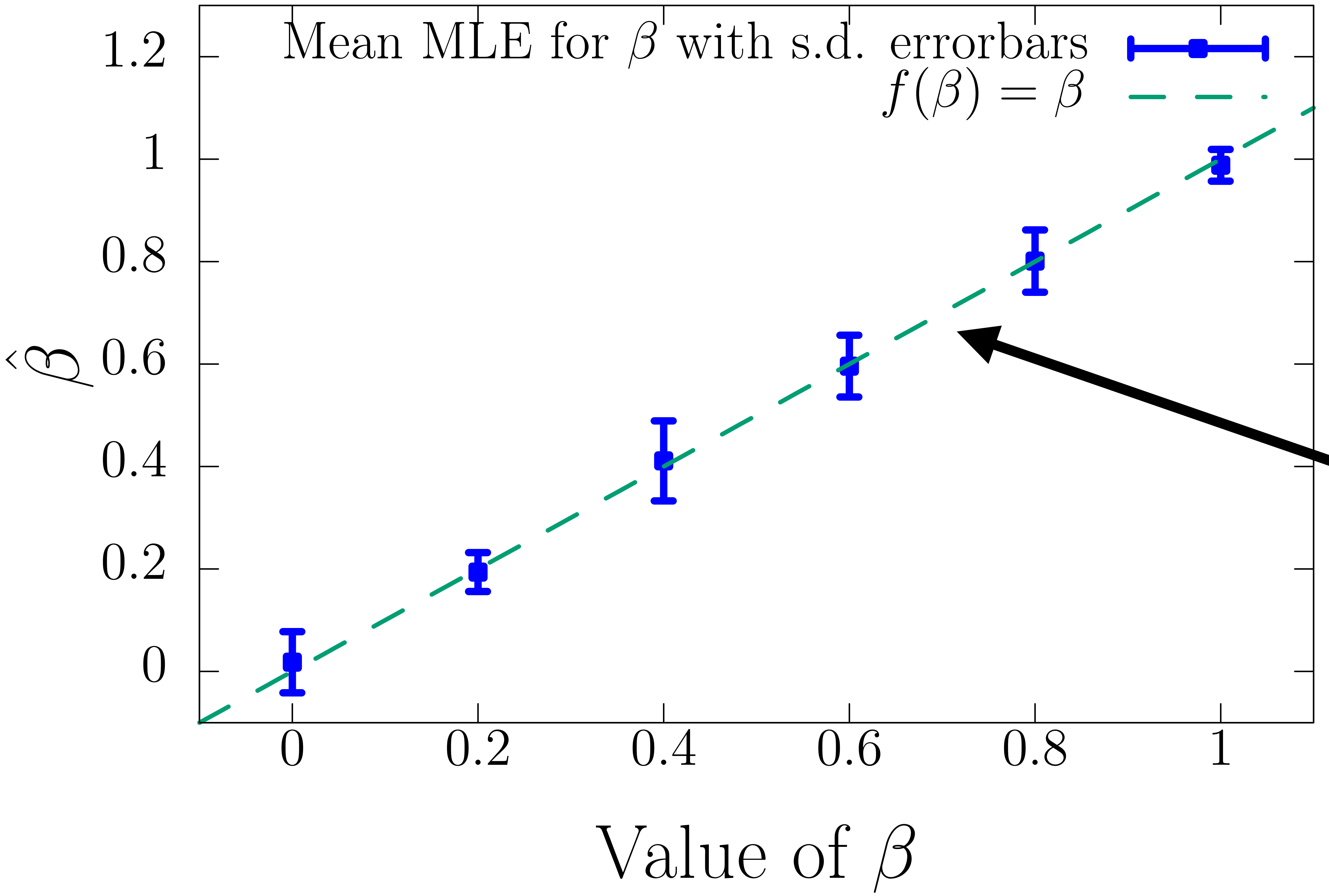
Distinguishing networks generated by BA vs RP



Test: given an artificial network grown that is generated from part BA and part RP, can we recover the proportion each model contributed using model likelihood?

Distinguishing networks generated by BA vs RP

So, we can distinguish between even very similar models



Aim: be close to green line

β

Barabasi-Albert **0** **1** **Rank-Preference**

Example: Modeling the AS topology

University of Oregon
Routeviews dataset

Candidate models to mix:

Positive-Feedback-Preference (PFP)

Nodes chosen uniformly at random

Example: Modeling the AS topology

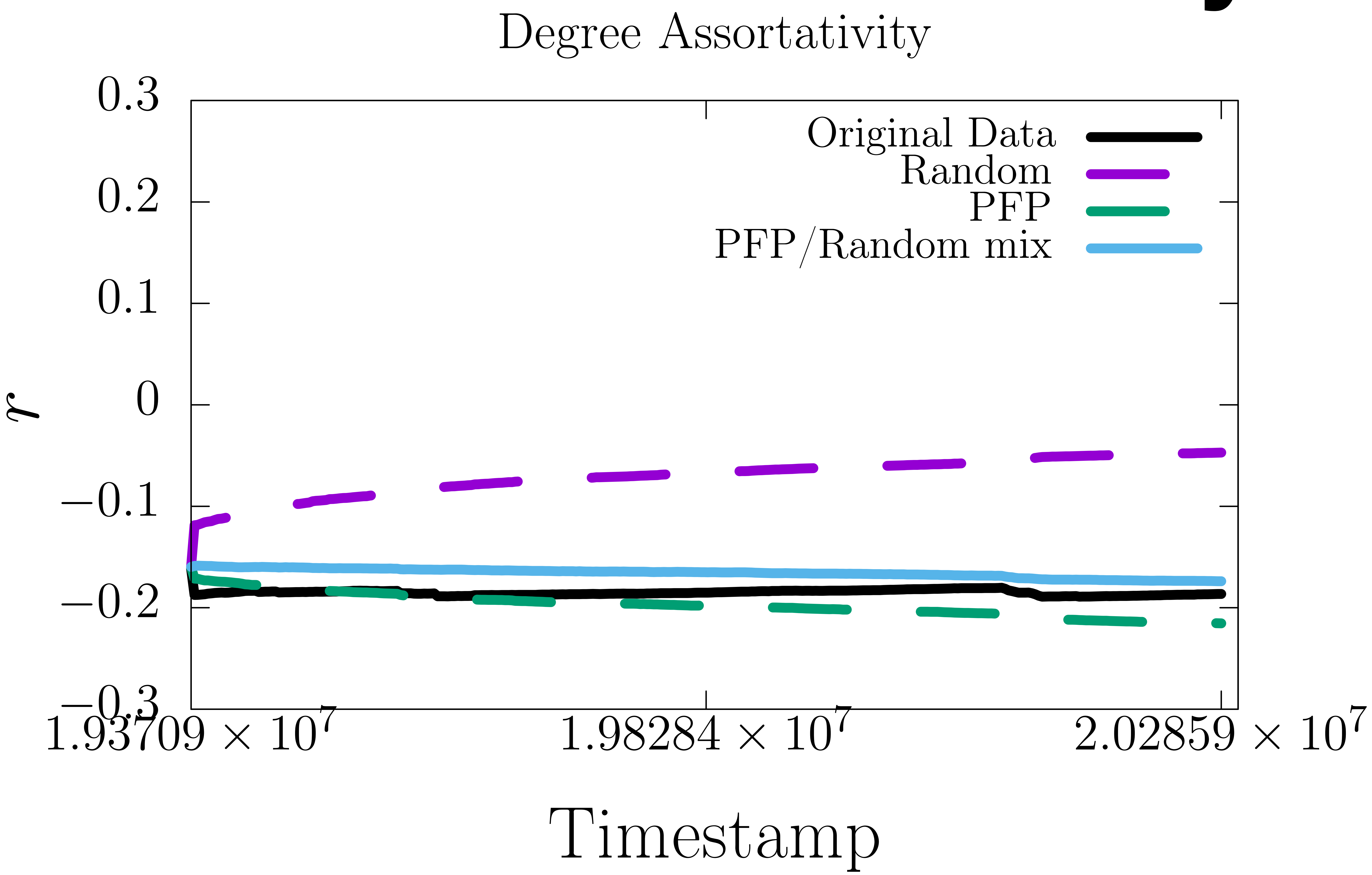
University of Oregon
Routeviews dataset

Candidate models to mix:

Positive-Feedback-Preference (PFP)

Nodes chosen uniformly at random

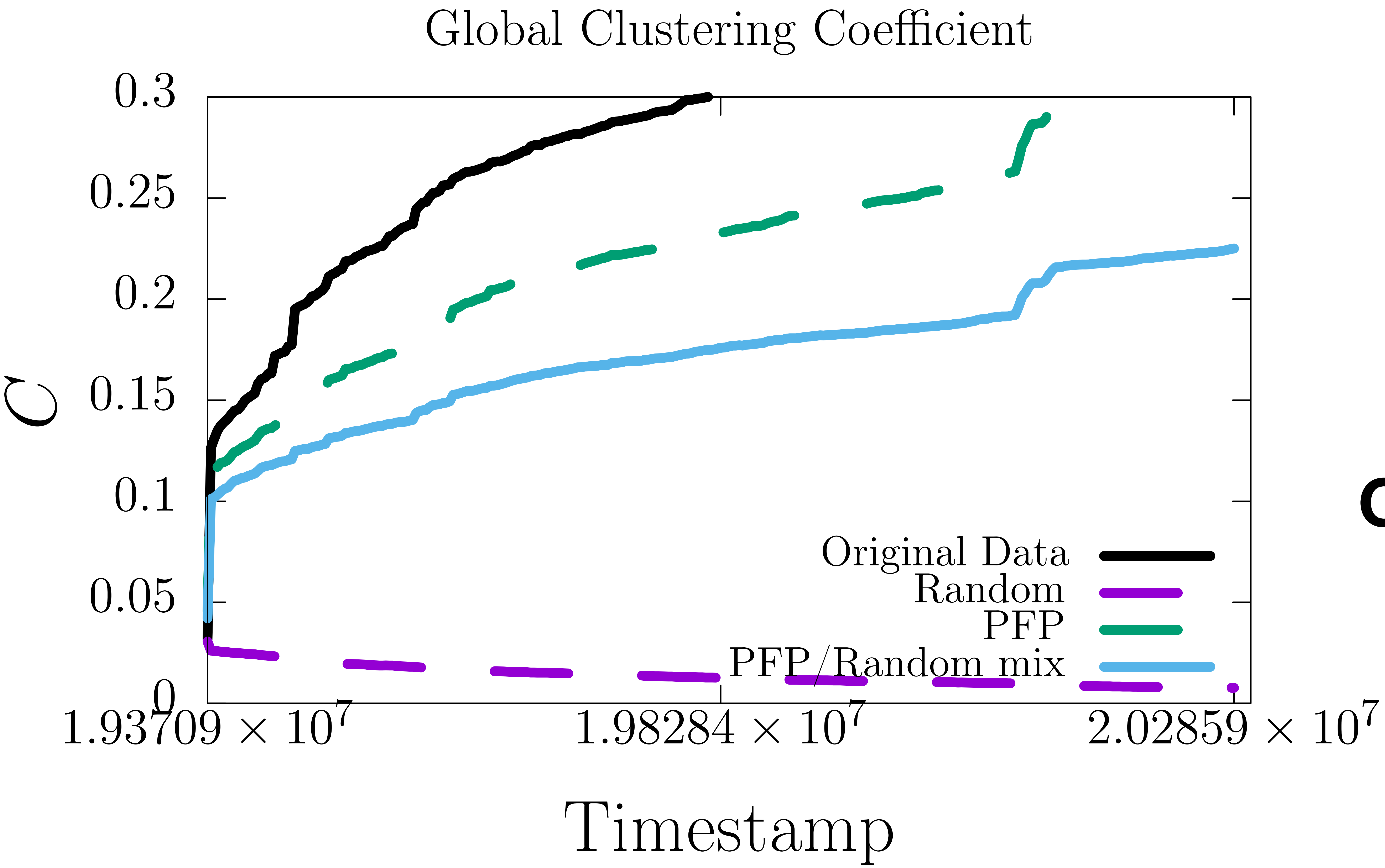
Takeaways



Some stats match well!

Degree Assortativity

Takeaways



Others not so well!

Clustering coefficient

Conclusion and Next Steps

Conclusion and Next Steps

Model likelihood as a tool for identifying the contribution of different network growth mechanisms - even similar-looking ones

Conclusion and Next Steps

Model likelihood as a tool for identifying the contribution of different network growth mechanisms - even similar-looking ones

Can be used to find the best fit of a mixture of models to real data that takes into account the whole network's evolution

Conclusion and Next Steps

Model likelihood as a tool for identifying the contribution of different network growth mechanisms - even similar-looking ones

Can be used to find the best fit of a mixture of models to real data that takes into account the whole network's evolution

Use of model likelihood to detect changepoints: drop in likelihood indicating network change.

Conclusion and Next Steps

Model likelihood as a tool for identifying the contribution of different network growth mechanisms - even similar-looking ones

Can be used to find the best fit of a mixture of models to real data that takes into account the whole network's evolution

Use of model likelihood to detect changepoints: drop in likelihood indicating network change.

Thanks for listening!

n.a.arnold@qmul.ac.uk