Mixed and time varying models for evolving complex networks

Naomi Arnold, Raul Mondragon, Richard Clegg

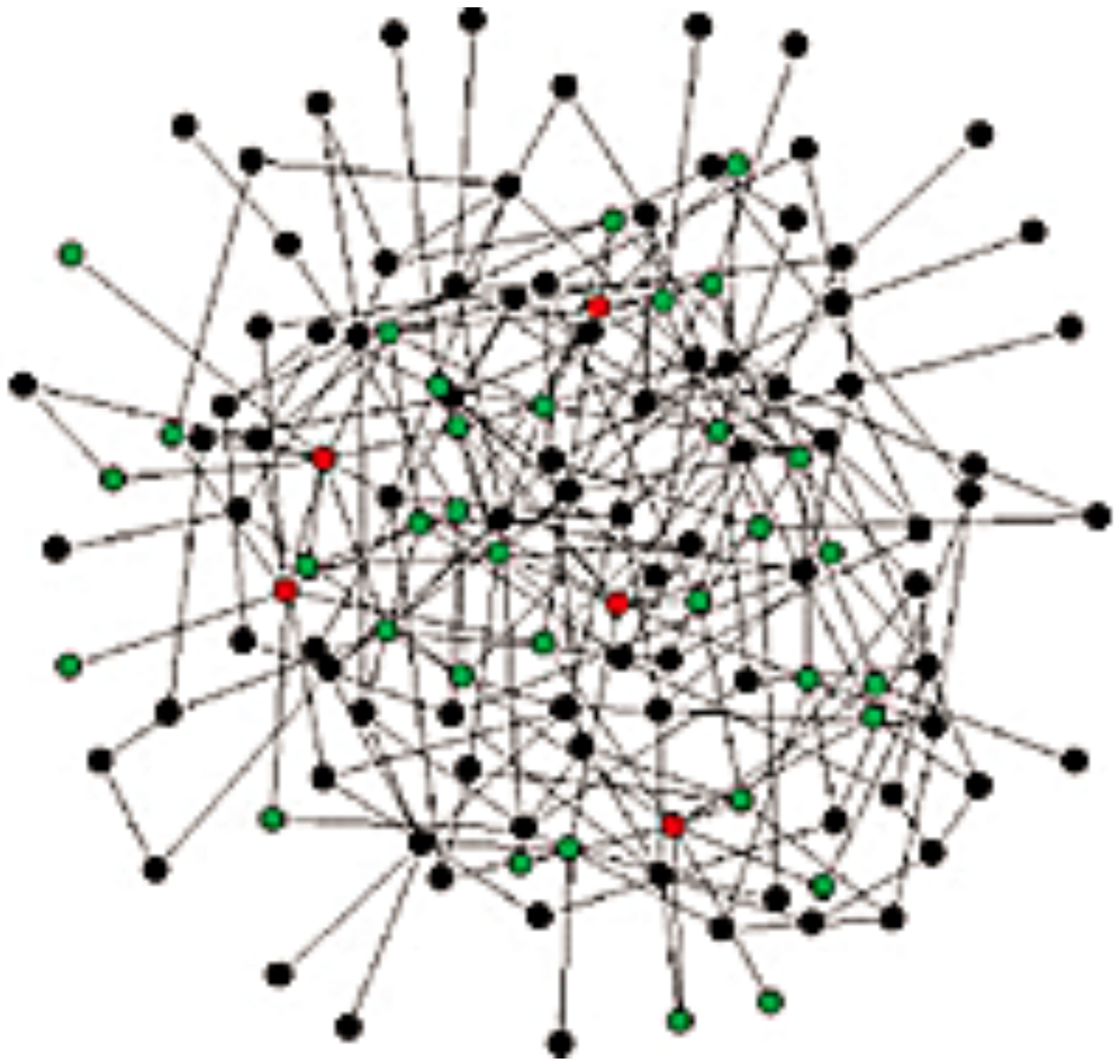
MoN 17 2018

Queen Mary
University of London

Outline

1. Background
2. Mixed models
3. Time varying models
4. Future directions

Generative models for network structure: Static



- Generates single network snapshots
- Examples: Erdős–Rényi, Stochastic Block Model, Configuration Model, ERGMs
- Community detection, centrality

Generative models for network structure: Dynamic

- Generates network by addition of nodes and links
- Investigate how network structural features emerge

Example: Barabási-Albert model

Characterised by:

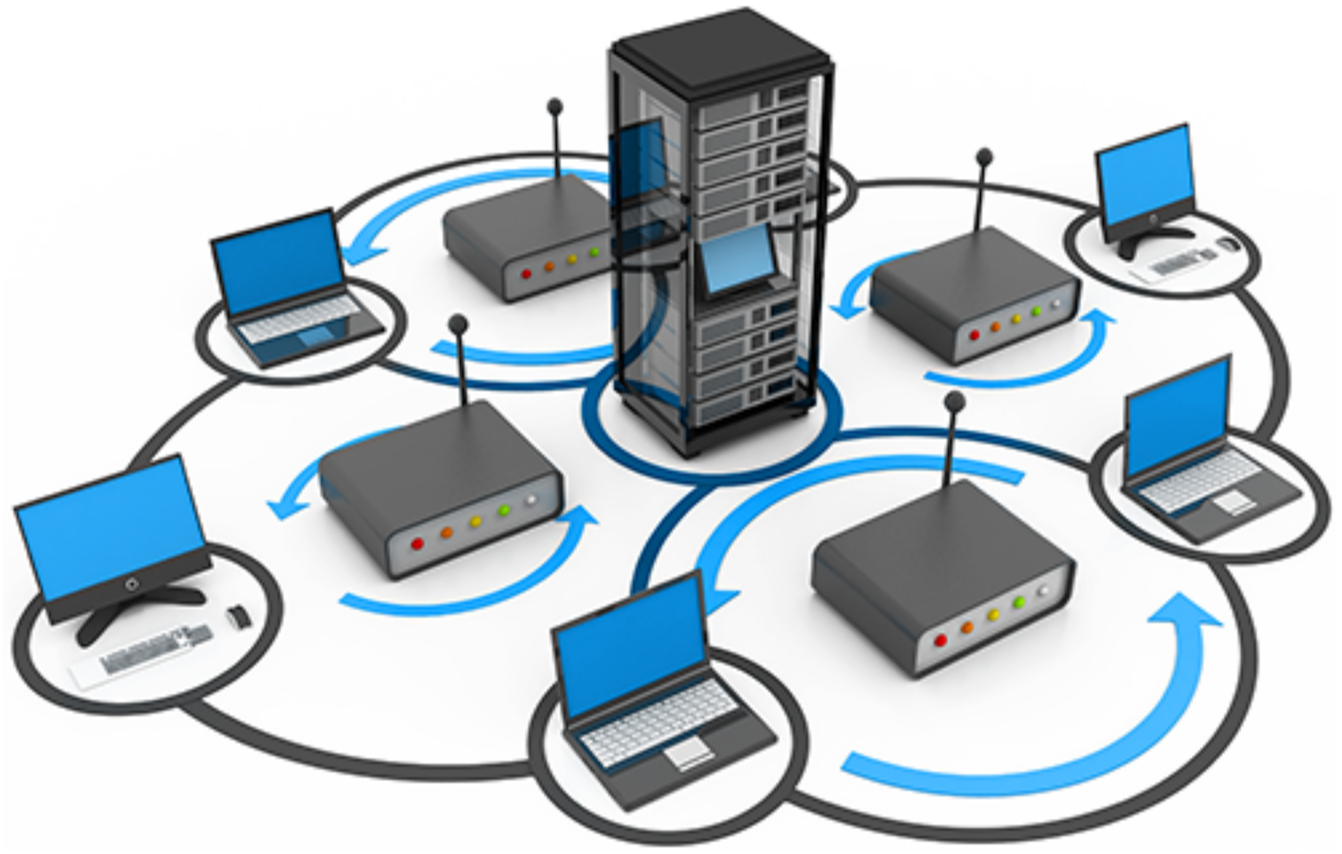
<u>Growth</u>: network grows by iterative addition of a single node with m neighbours

<u>Preferential attachment:</u> new node connects to existing node i with probability $\propto k_i$.

i.e. better connected nodes likelier to attract new links

$m = 2$

Examples of evolving networks



Usual modelling assumptions

- Network grows using a single constant mechanism
- Good for deriving theoretical results…
- … but not great for making inferences from real data

Our aim

To relax the usual modelling assumptions made, to better comprehend a model governing a network’s evolution

“Single mechanism” → more than one mechanism, to uncover the roles of each of them.

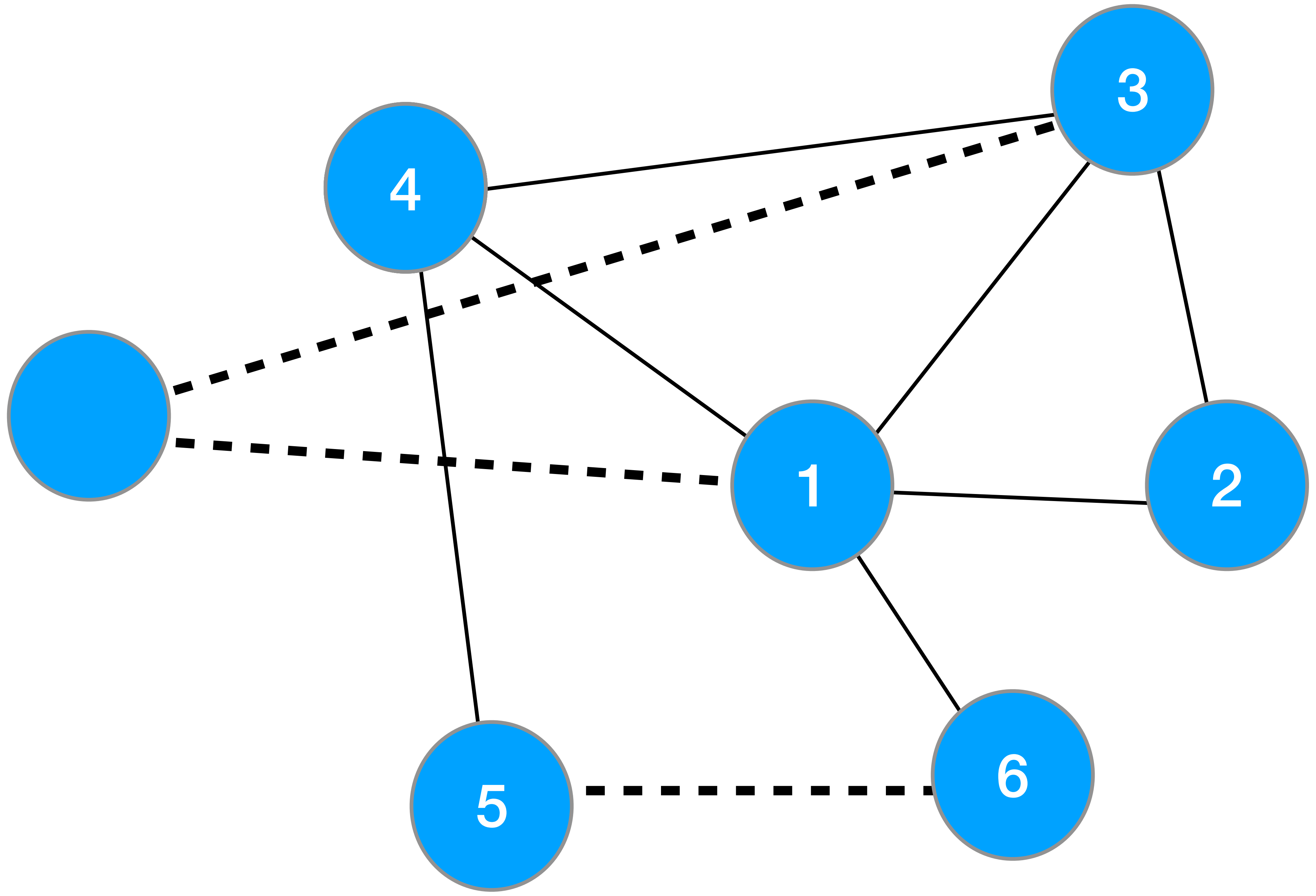
“Constant” → changing in time, to understand how these roles may change over time

Model for evolving networks

Seed network

Attachment probabilities

Action (new node/internal link)



$$p_i$$

corresponding to

$\mathbb{P}($**choose node** $i)$

Attachment probabilities

$$p_i \propto 1$$

Random/neutral model. All nodes equally likely

$$p_i \propto f(k_i)$$

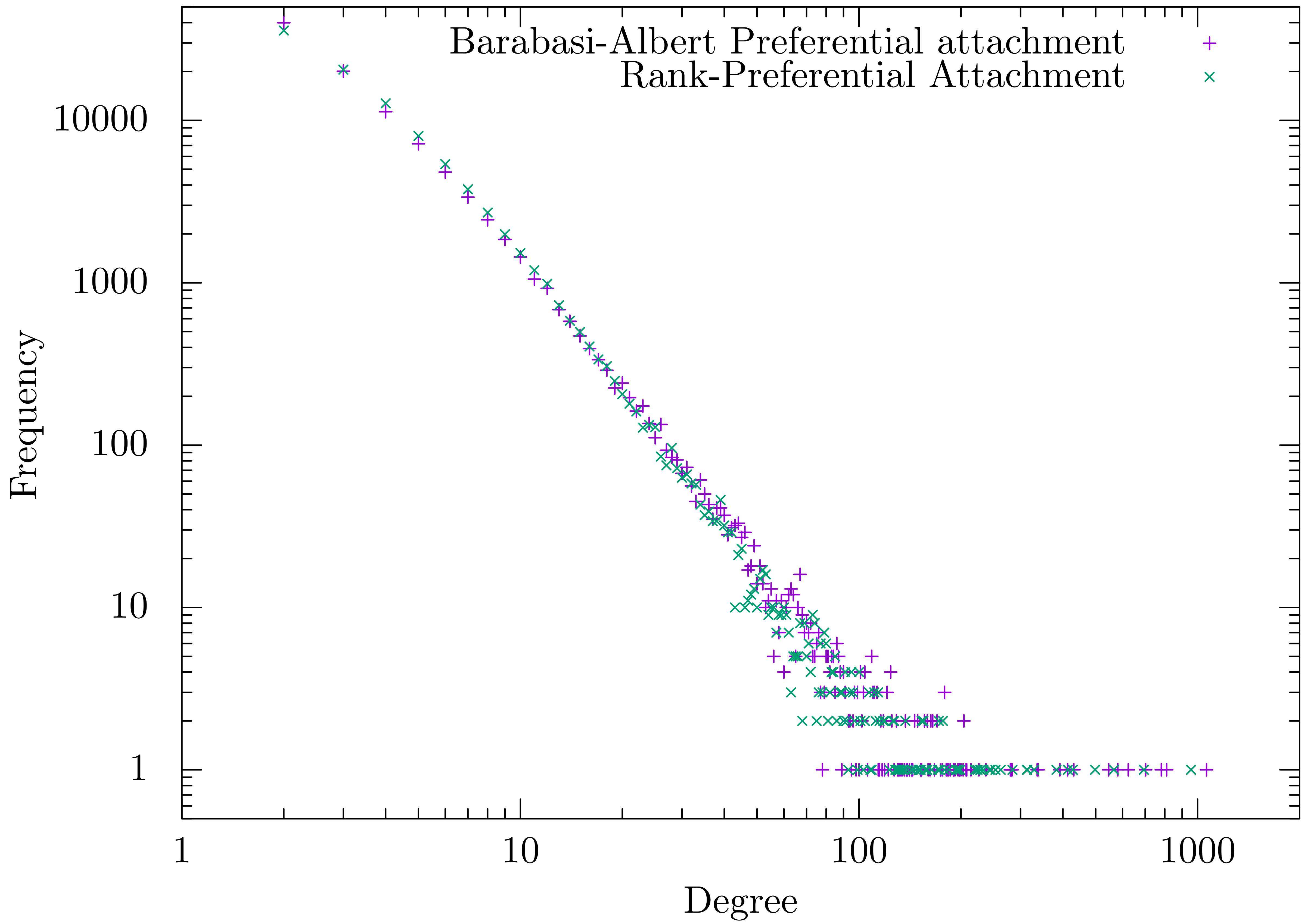
Function of node i's degree, e.g. BA model

$$p_i \propto f(\eta_i)$$

Function of some other intrinsic node property

Problem: How do we quantify how good a fit a model is to real data?

Moving away from descriptive statistics



Barabási Albert model

$$p_i \propto k_i$$

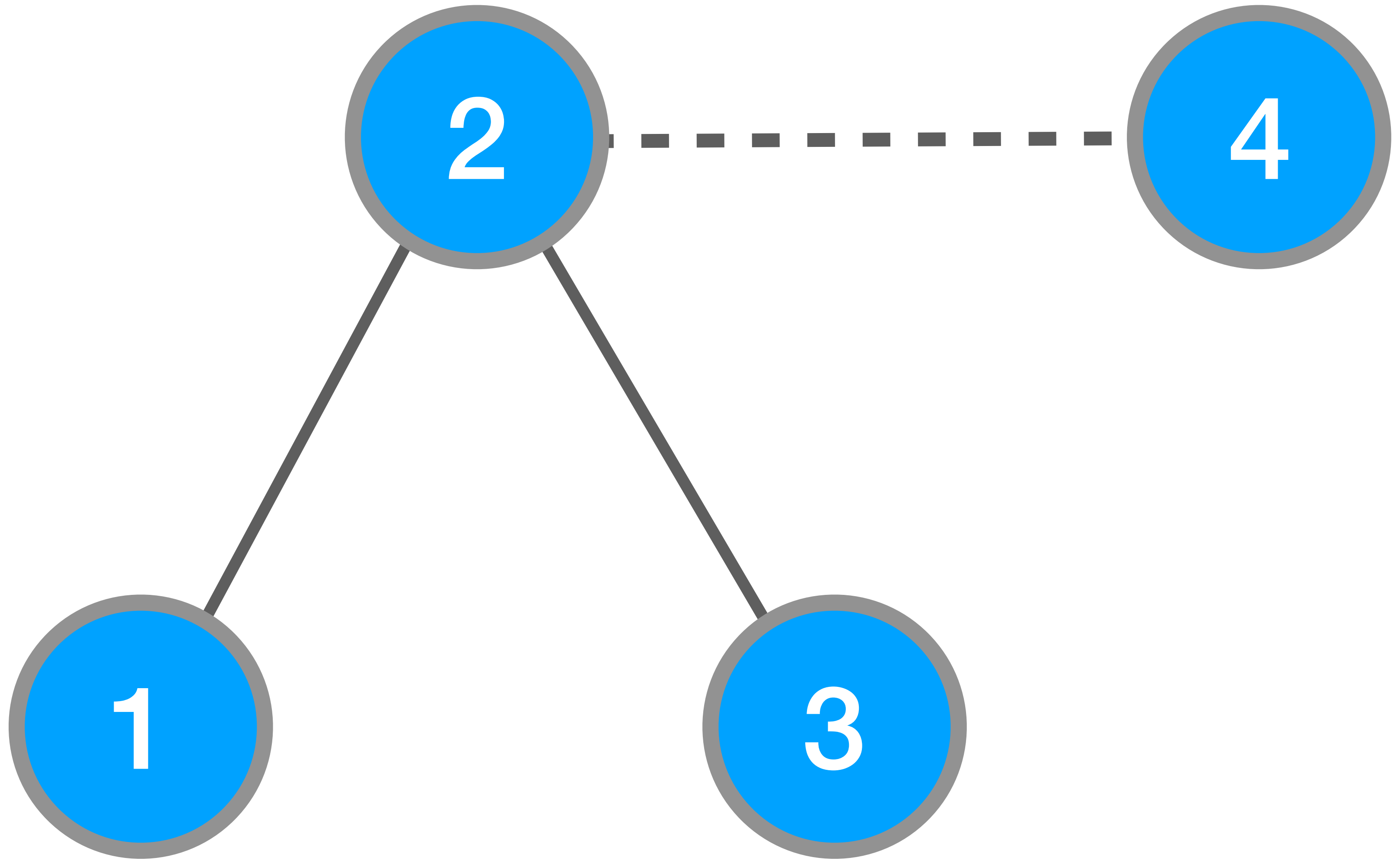
Static rank-preference model

$$p_i \propto i^{-\alpha}$$

Problem: different models may generate networks with similar descriptive statistics!

Different approach: Model likelihood

Likelihood of model given observation = probability of seeing observation given model



Likelihood of random/uniform model given by:

$$\mathbb{P}_{rand}(\textbf{Choose node } 2) = \frac{1}{3}$$

Likelihood of BA preferential attachment model given by:

$$\mathbb{P}_{BA}(\textbf{Choose node } 2) = \frac{2}{1+2+1} = \frac{1}{2}$$

[R. Clegg, B. Parker, M. Rio 2016: Likelihood-based assessment of dynamic network models]

Likelihood of model given observed period of network's evolution

Network (random variable): $G := G_t$

Observations (snapshots): $\mathbf{g} = \left(g_1, g_2, \ldots, g_n\right)$

Model possibly with parameters: $M(\theta)$

$$L(M(\theta)\,|\,\mathbf{g}) = \prod_{t=1}^{n} \mathbb{P}(G_t = g_t\,|\,G_{t-1} = g_{t-1}, \ldots, G_1 = g_1, M(\theta))$$

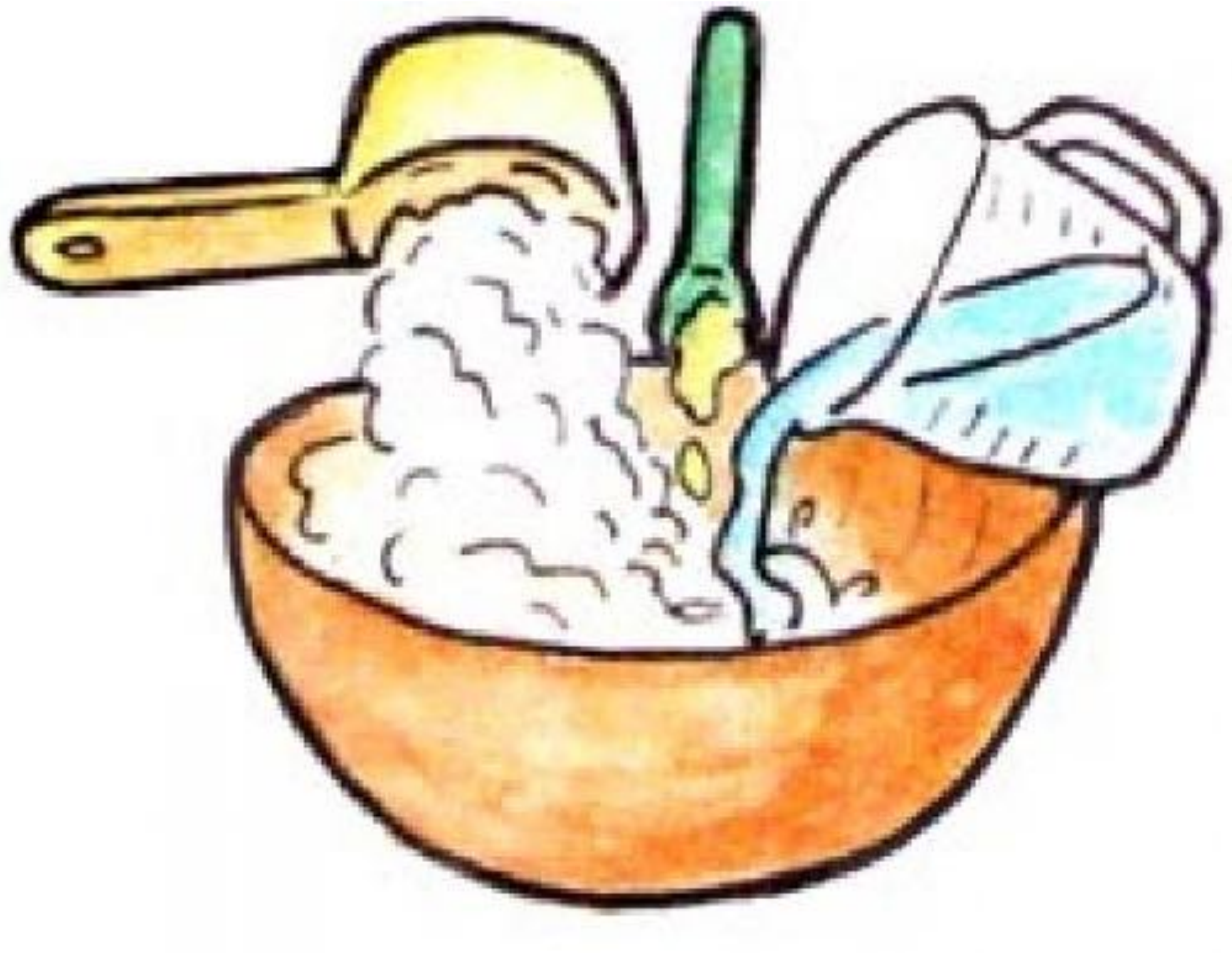
[R. Clegg, B. Parker, M. Rio 2016: Likelihood-based assessment of dynamic network models]

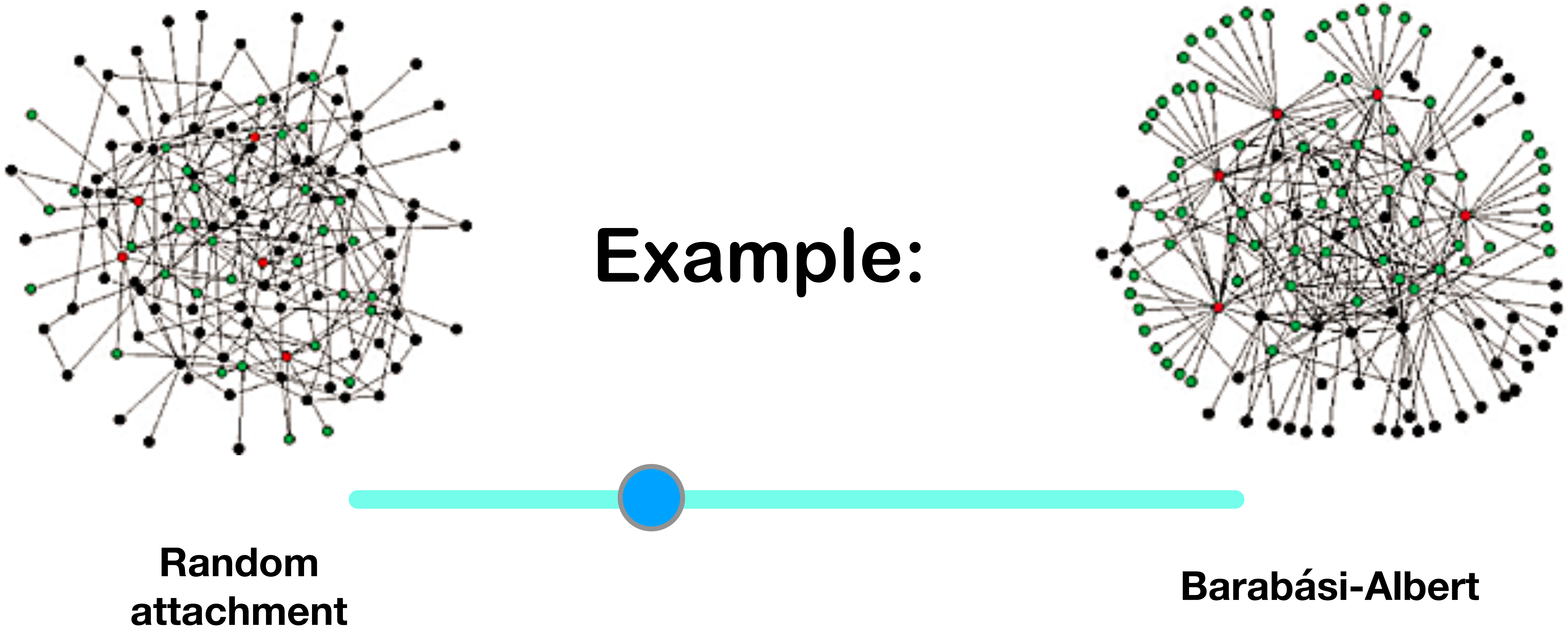
Likelihood: Remarks

- Quickly calculated, compared to generating networks
- Given a number of models, can define the 'best' as that which has the highest likelihood
- For models with parameters, can find maximum likelihood estimators for params

Assumption relaxation 1: Combining attachment probabilities

Idea: attachment in networks likely to be driven by a mixture of factors





This example explored in [Ghoshal, Chi, Barabási 2014: Uncovering the role of elementary processes in network evolution]

More formally:

$$p_i = \sum_l \beta_l p_i^{(l)} \quad \text{with} \quad \sum_l \beta_l = 1$$

a weighted combination of models

$p_i^{(l)}$ probability of choosing node i according to model l

[R. Clegg, B. Parker, M. Rio 2016: Likelihood-based assessment of dynamic network models]

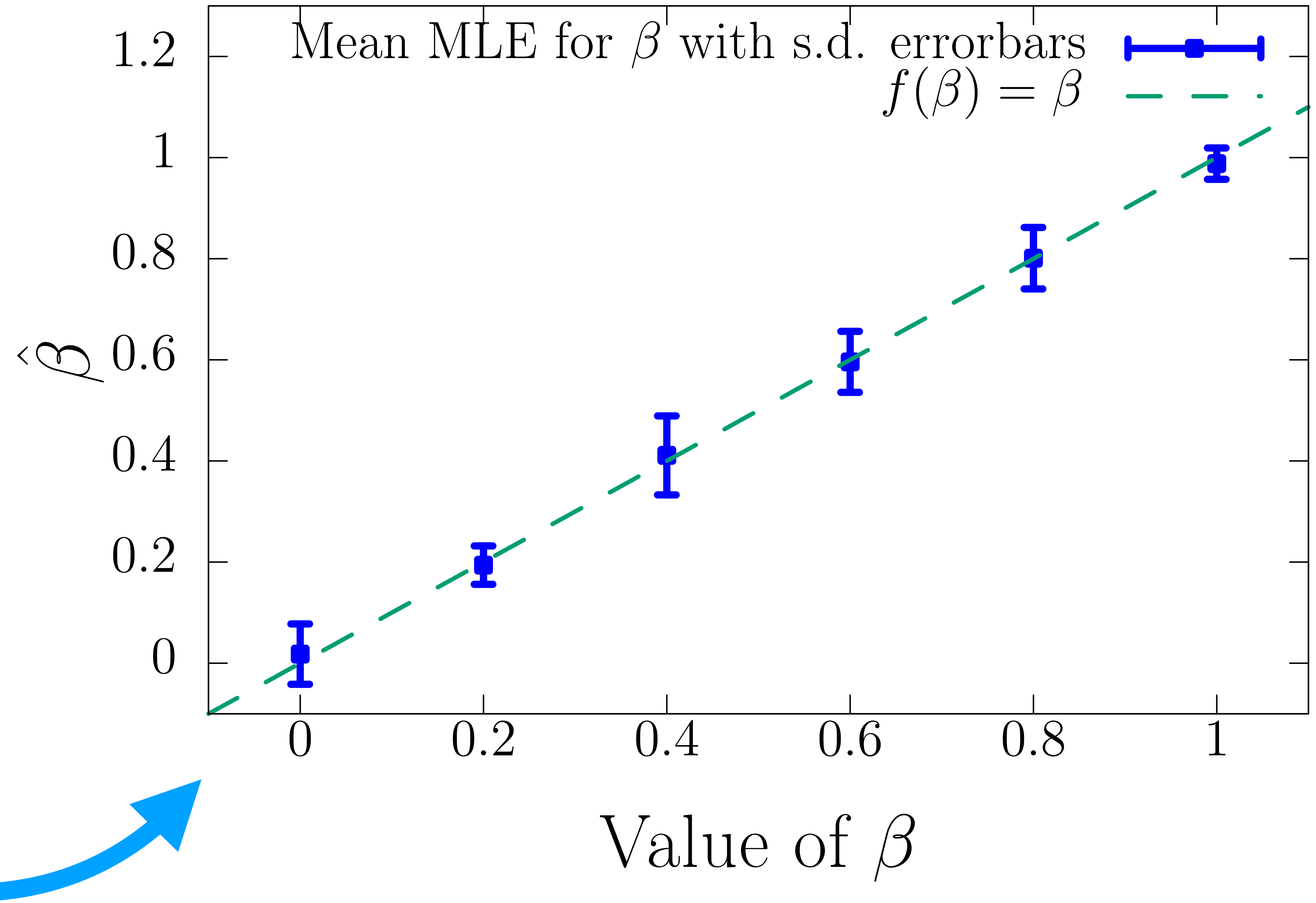
Can address the earlier problem with descriptive statistics using likelihood measure!

Distinguishing using maximum likelihood estimation

Having generated artificial networks using:

$$p_i = \beta p_i^{RP} + (1-\beta) p_i^{BA}$$

We can accurately recover the proportion β **as an MLE!**

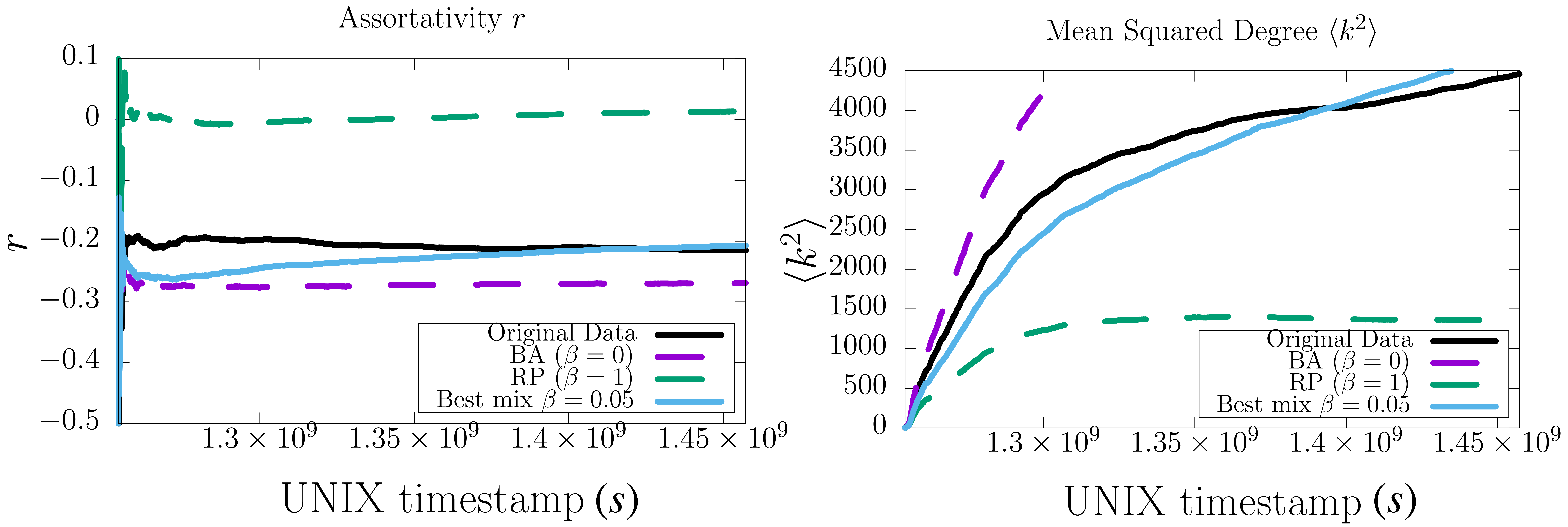


Real data example: Math Overflow Social Network

- Q&A site for mathematicians with problems
- Nodes are users
- An undirected edge between node A and B if A answers a question by B, A comments on B’s answer or question
- Multiple edges collapsed
- Models components tested: BA, static rank preference

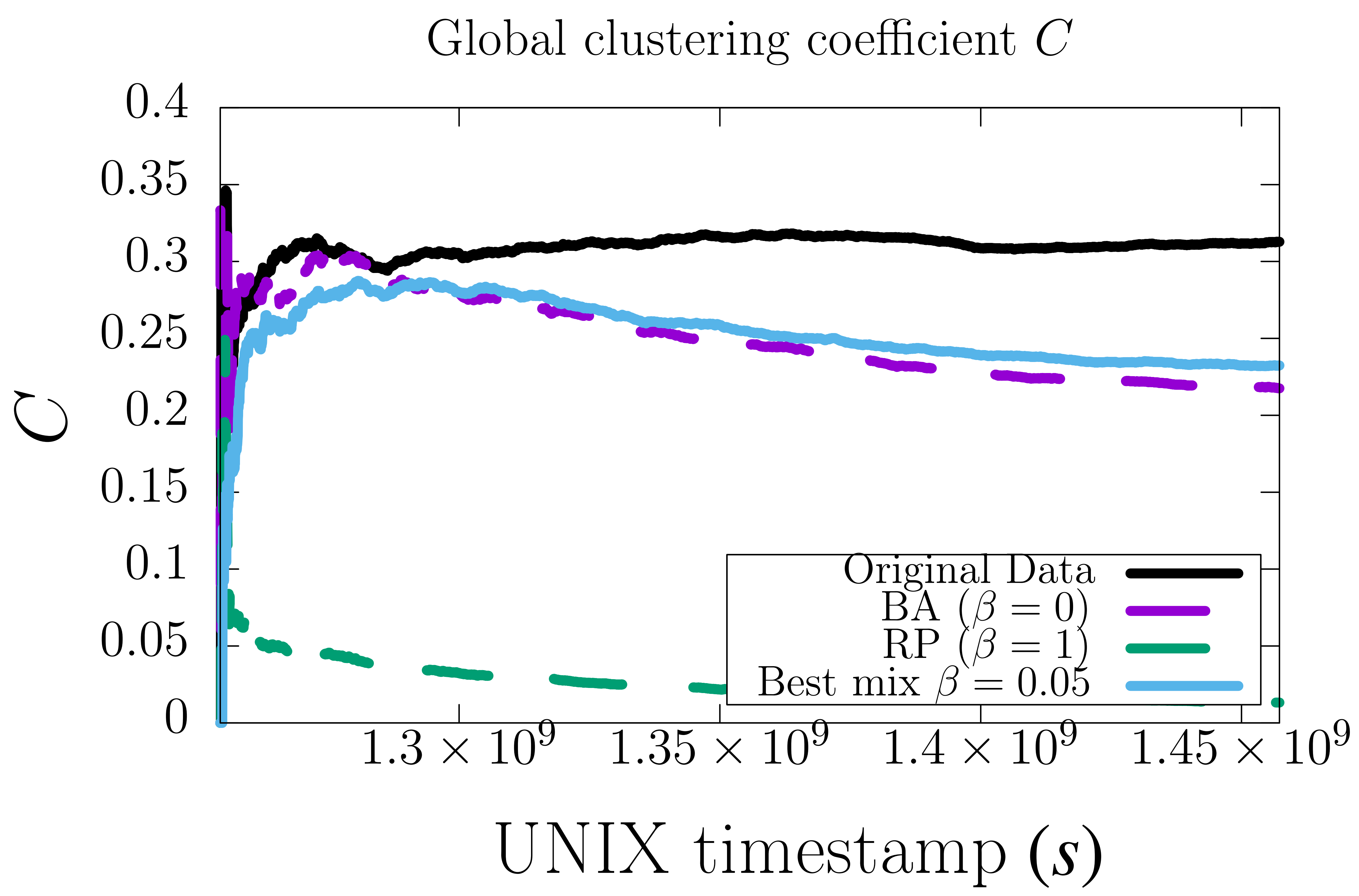
Dataset from [A. Paranjape, A. Benson, J. Leskovec 2017: Motifs in temporal networks]

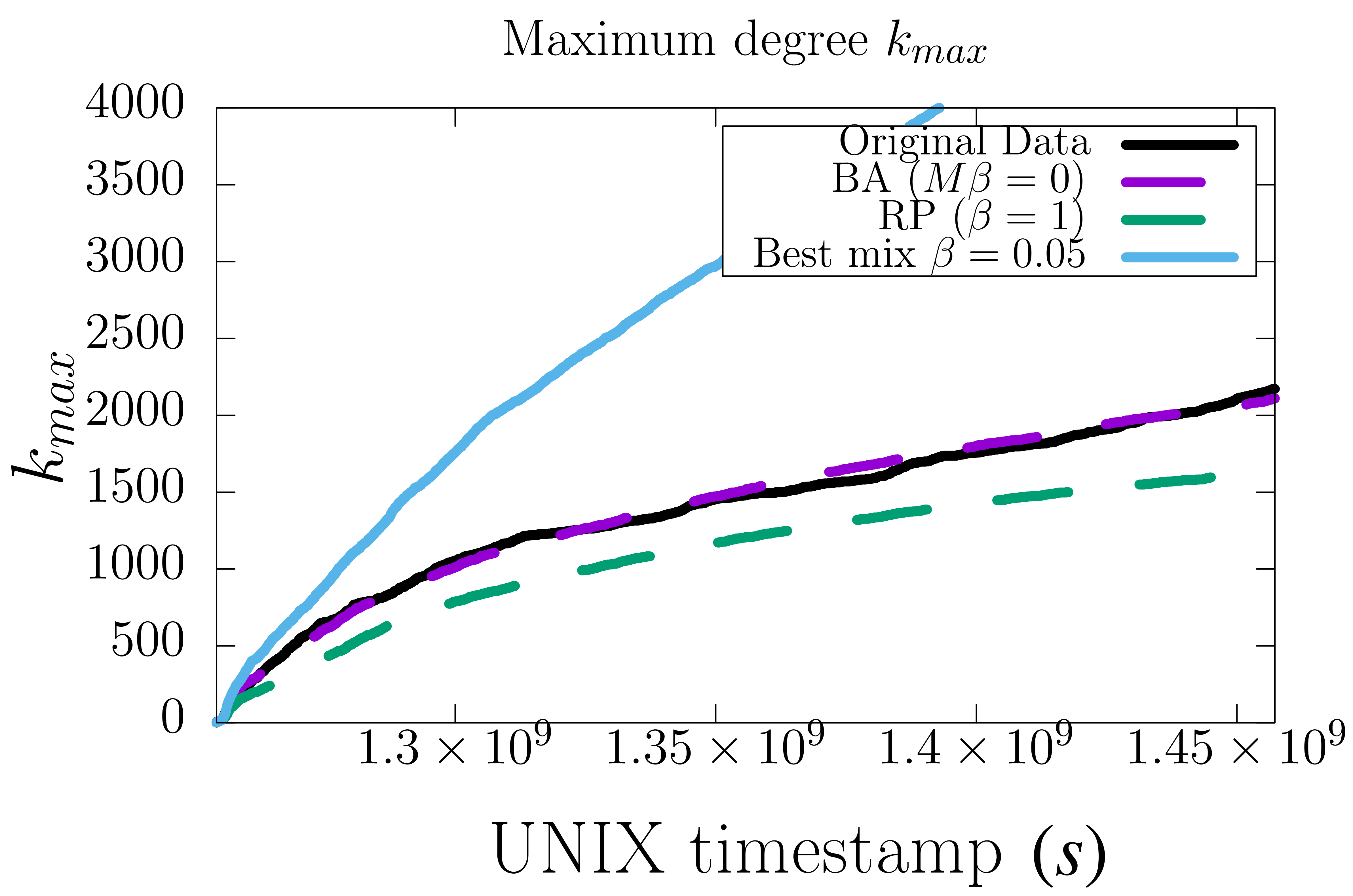
Real data example: Math Overflow Social Network



Dataset from [A. Paranjape, A. Benson, J. Leskovec 2017: Motifs in temporal networks]

Real data example: Math Overflow Social Network



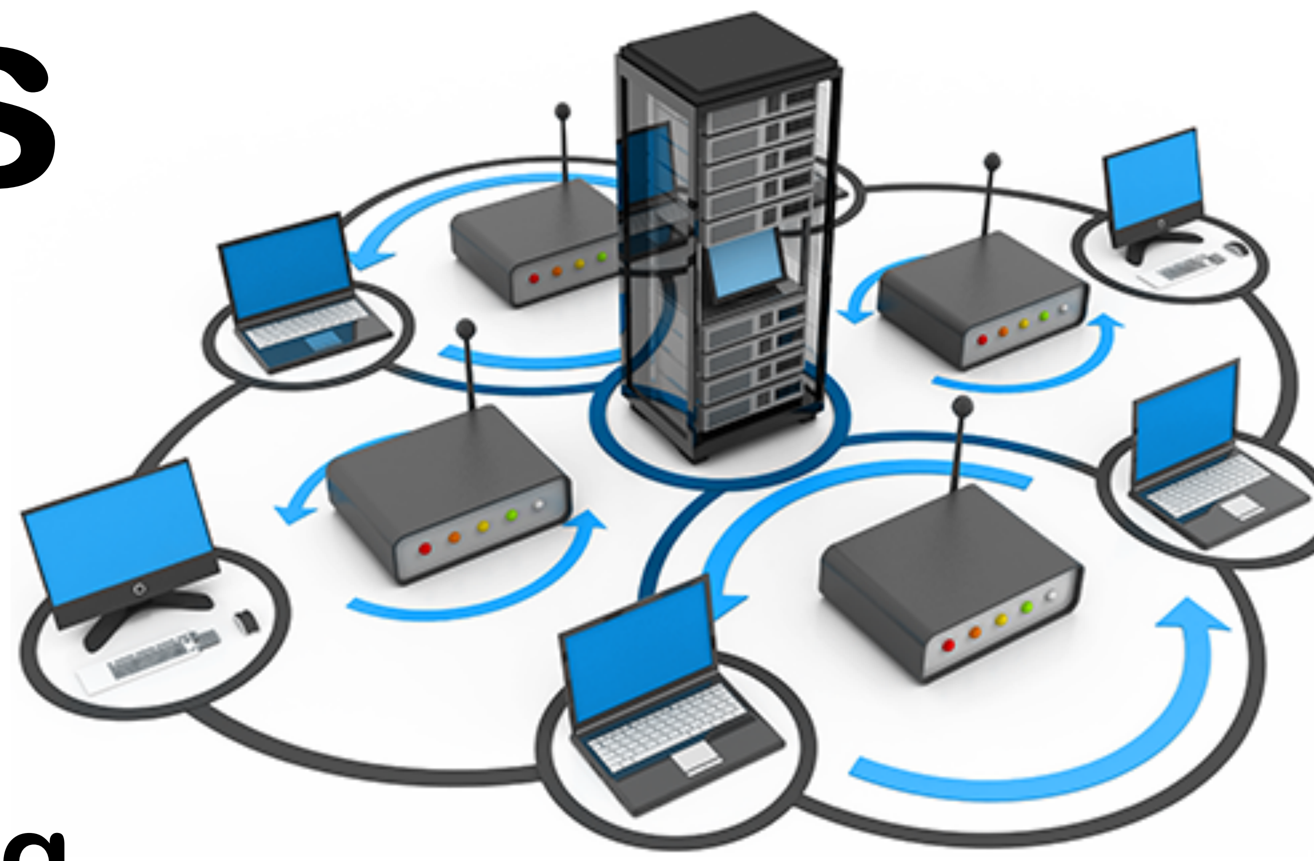


Dataset from [A. Paranjape, A. Benson, J. Leskovec 2017: Motifs in temporal networks]

Remarks

- Mixed model often generates network with better match on stats than single model components alone
- Process only gives the highest likelihood mix of the components tested - not guaranteed to be a good model.
- Searching through parameter space becomes expensive with more than two model components. Candidate problem for ML techniques.
- Work in progress: applying simulated annealing to model fitting

Assumption relaxation 2: Time varying models



In case you missed it

NHS Million and 8 others follow

People you may know

Adorable Retrievers 🐶 and 2 others follow

Puppies 🐶 @PopularPups · 13h

- Networks may have changing growth regimes over time
- Can these changes be reflected in our modelling of networks?

Time varying model

For a set of models M_j which, given a network topology, assign probabilities to each node, introduce time varying model:

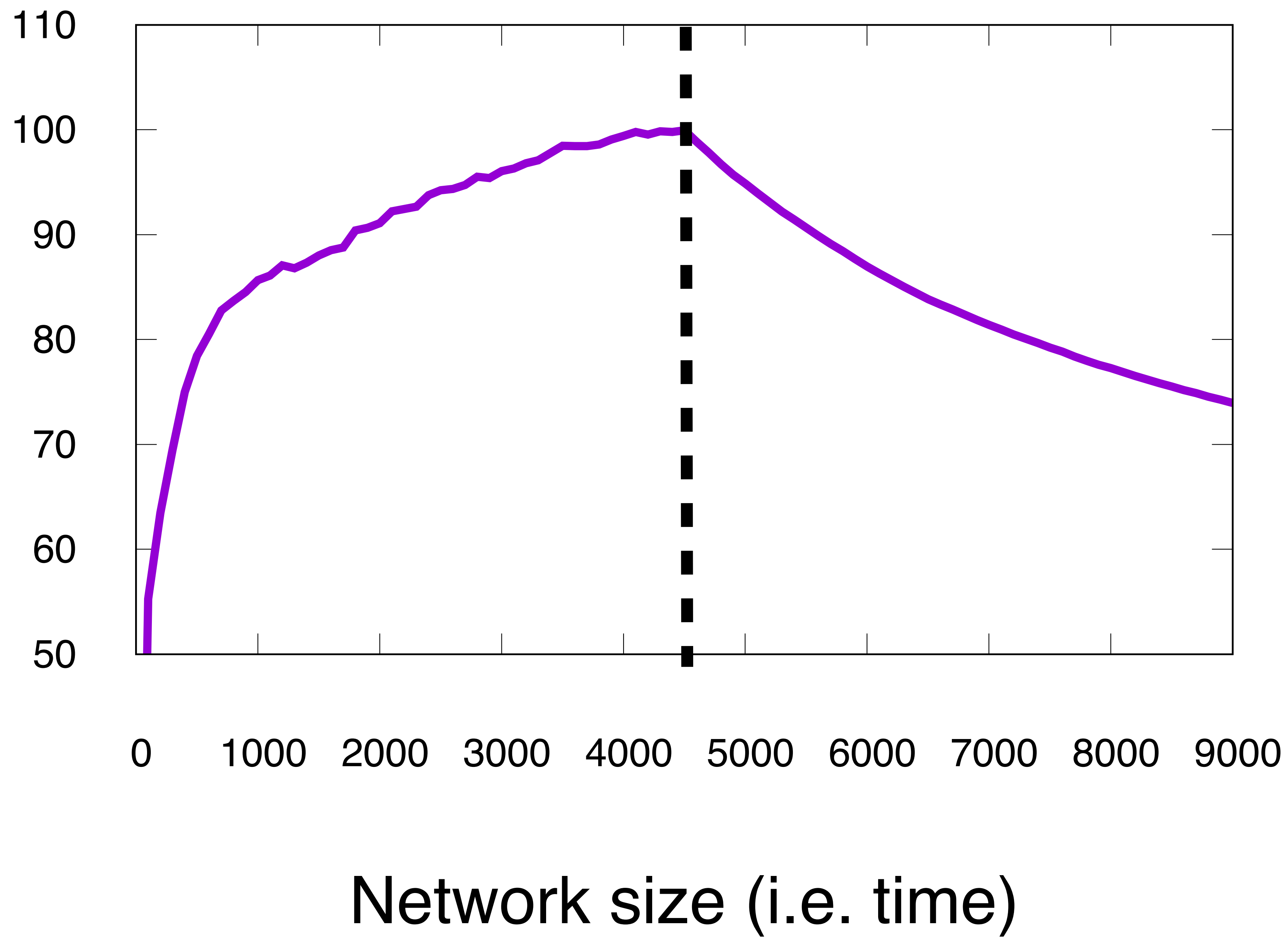
$$M(t) = \begin{cases} M_0 & t \in [t_0, t_1) \\ M_1 & t \in [t_1, t_2) \\ \vdots & \vdots \\ M_k & t \in [t_k, t_{k+1}) \end{cases}$$

Are such changes in a network's growth reflected in measurements we take of it?

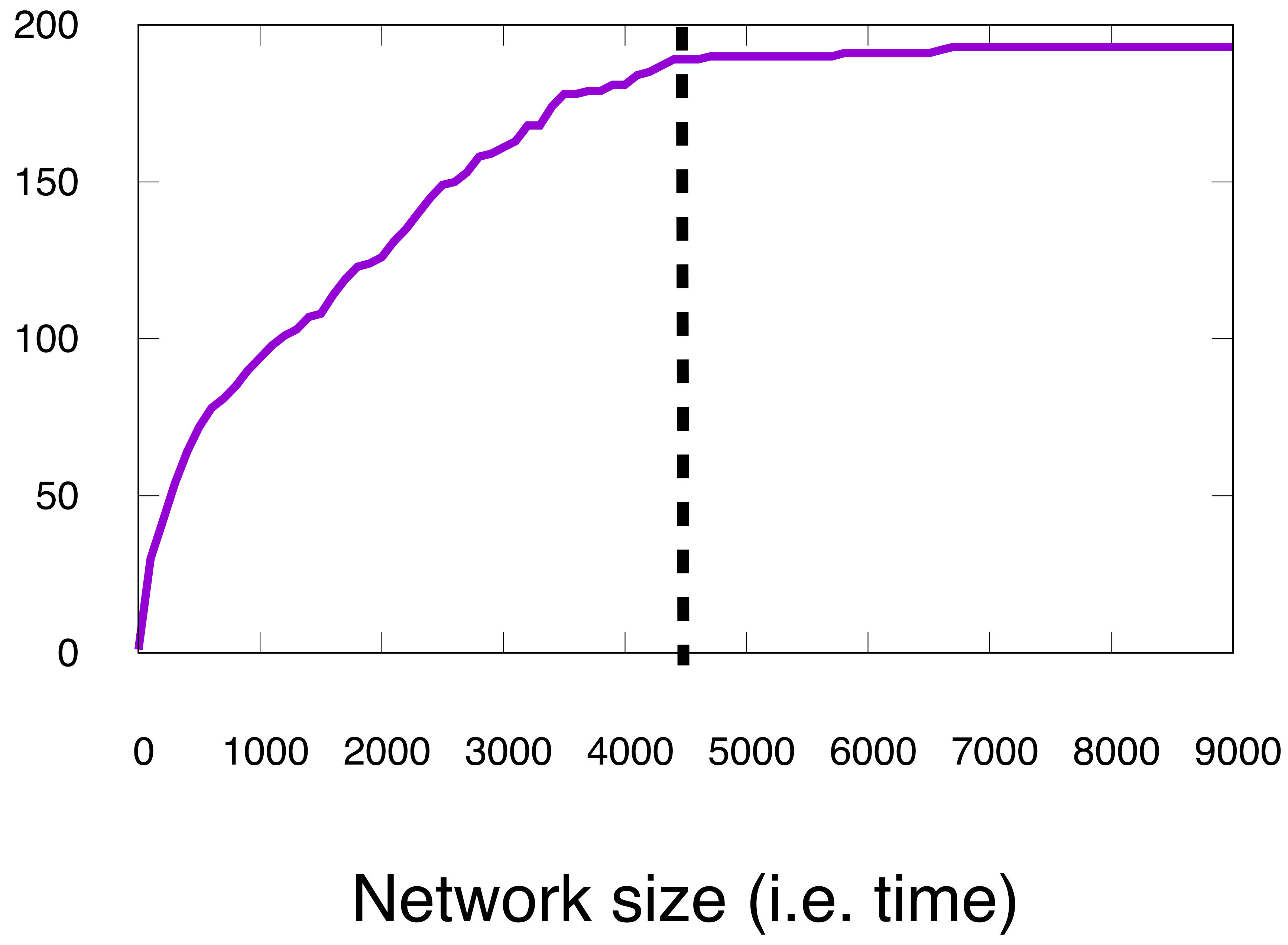
Sometimes yes

(Network generated with first half BA, second half neutral probabilities)

Mean squared degree $\langle k^2 \rangle$



Maximum degree k_{max}

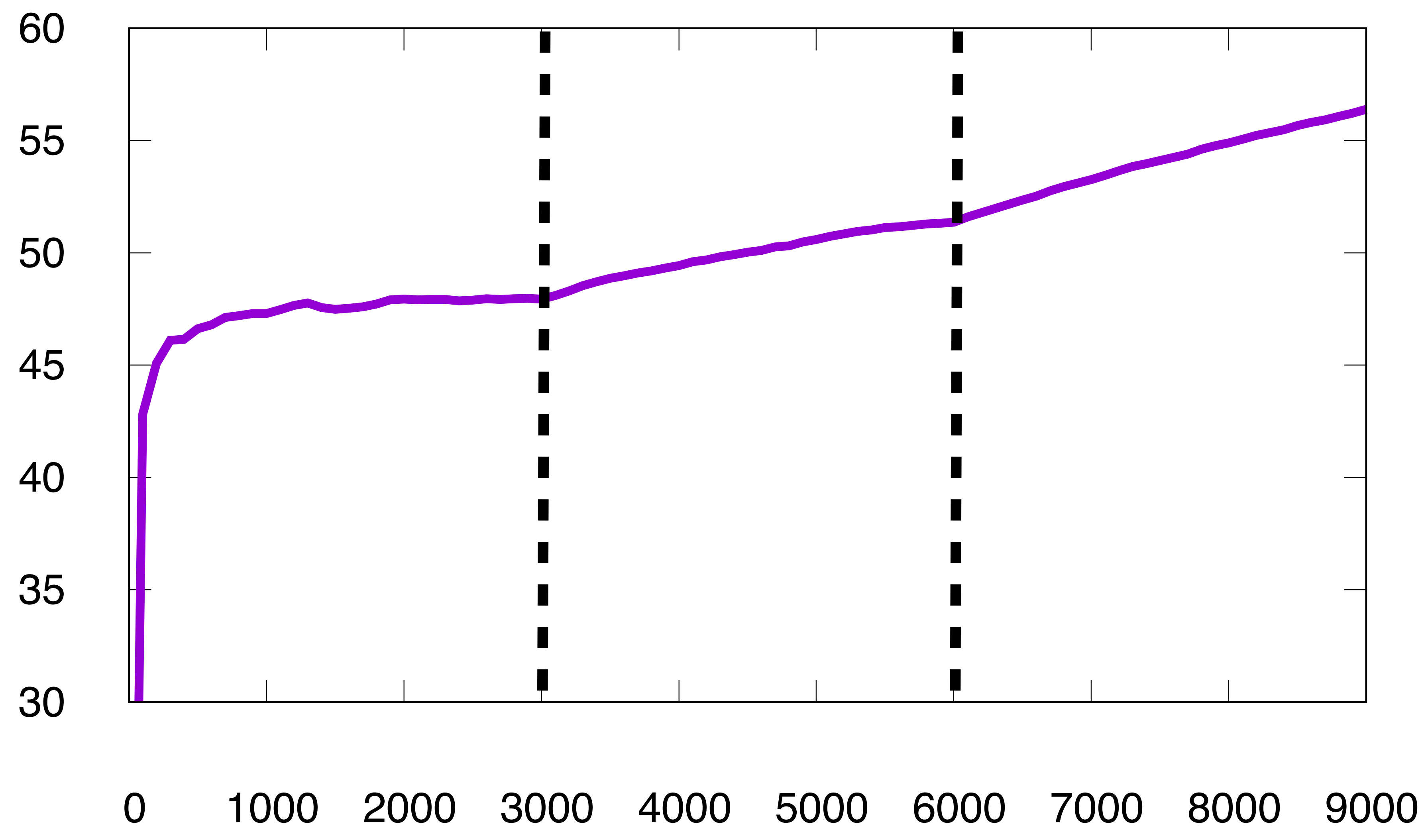


Are such changes in a network's growth reflected in measurements we take of it?

Other times less so...

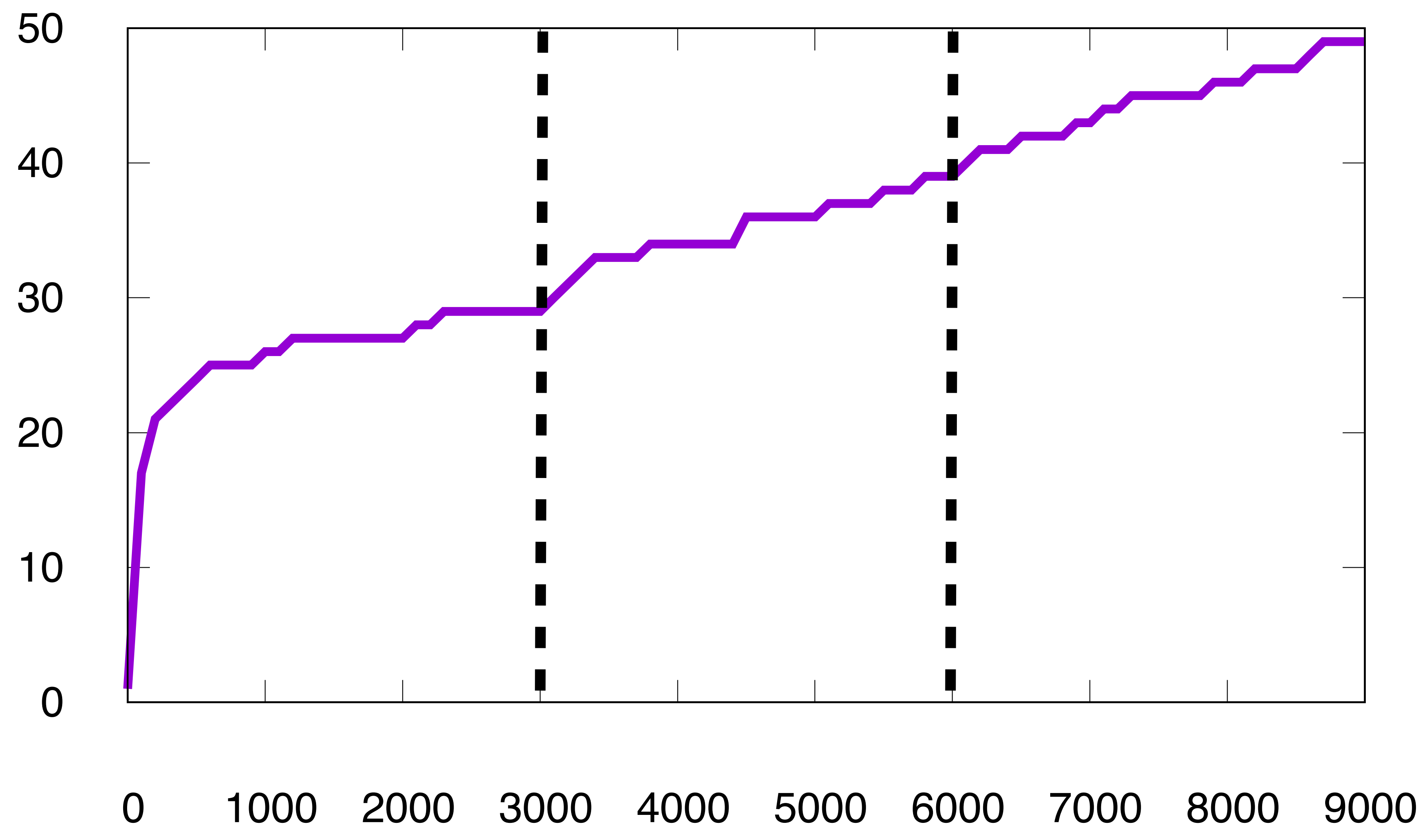
(Network first third: random, second: half random half BA, third: BA)

Mean squared degree $\langle k^2 \rangle$



Network size (i.e. time)

Maximum degree k_{max}

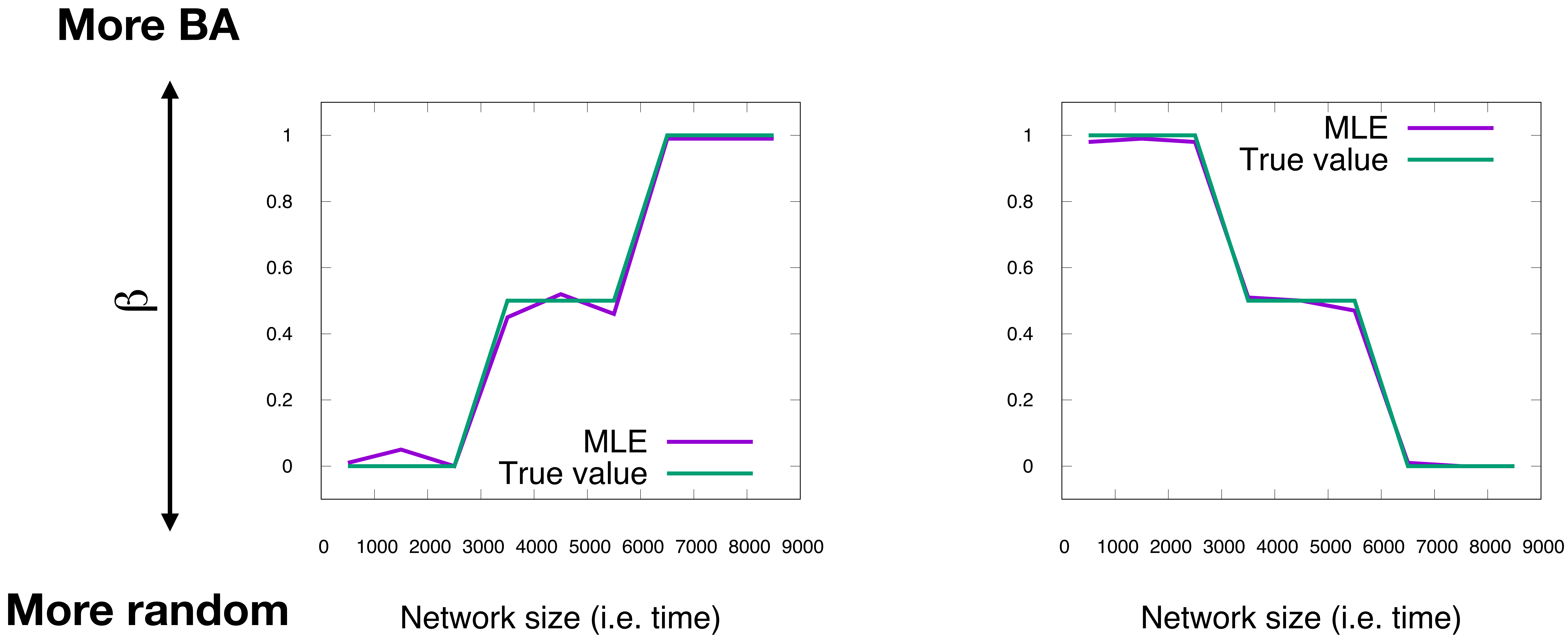


Network size (i.e. time)

Investigation with artificially generated networks

- Generate networks with different growth regimes
- What does this look like when we look at descriptive statistics?
- Can we detect these changes using likelihood-based techniques?

Tests on generated networks



Conclusions and future directions

- Mixed models allow us to uncover the different attachment mechanisms at play in a network's growth
- Fitting to different time periods may reveal how these mechanisms can change over time
- Currently developing a likelihood measure aimed at detection of changes in a network's growth regime
- Working on use with real data

Thank you for listening!
What are your questions?

github.com/narnolddd n.a.arnold@qmul.ac.uk @narnolddd