Imperial College London

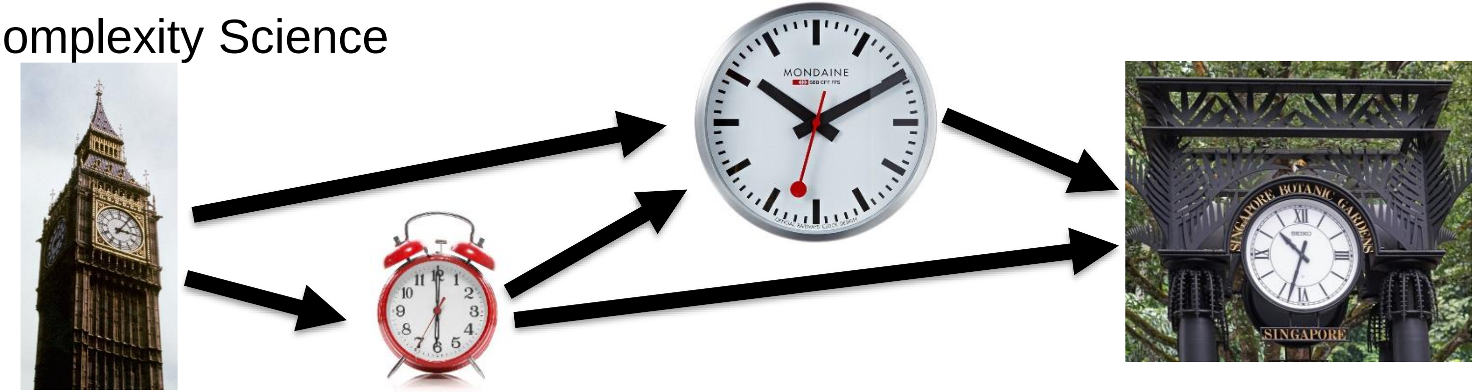
Slides at http://bit.ly/2oR62sI DOI: 10.6084/m9.figshare.9933806
CCS Singapore
4/10/2019
arXiv:1903.03667

The Longest Path in the Price Model

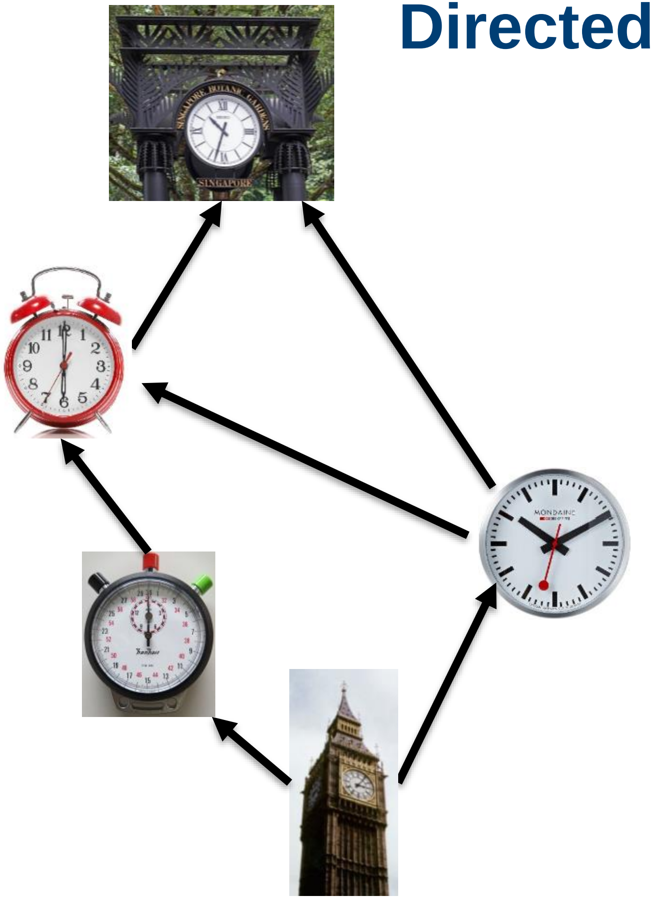
Complexity Science

TIM EVANS work with **Lucille Calmon, Vaiva Vasiliauskaite**

Centre for
Complexity Science



Directed Acyclic Graphs = DAG



- A directed network with no cycles
- Defines a ***Partial Order*** on set of nodes
  Order constrains direction of edges

e.g. **Temporal Vertex Networks**,
vertices assigned a time,
edges respect the ***arrow-of-time***

- Citation networks
  e.g. papers, patents, court judgements, blogs
- Task scheduling
- Food webs
- Cryptocurrency Transactions (e.g. IOTA)
- Causal set approach to quantum gravity

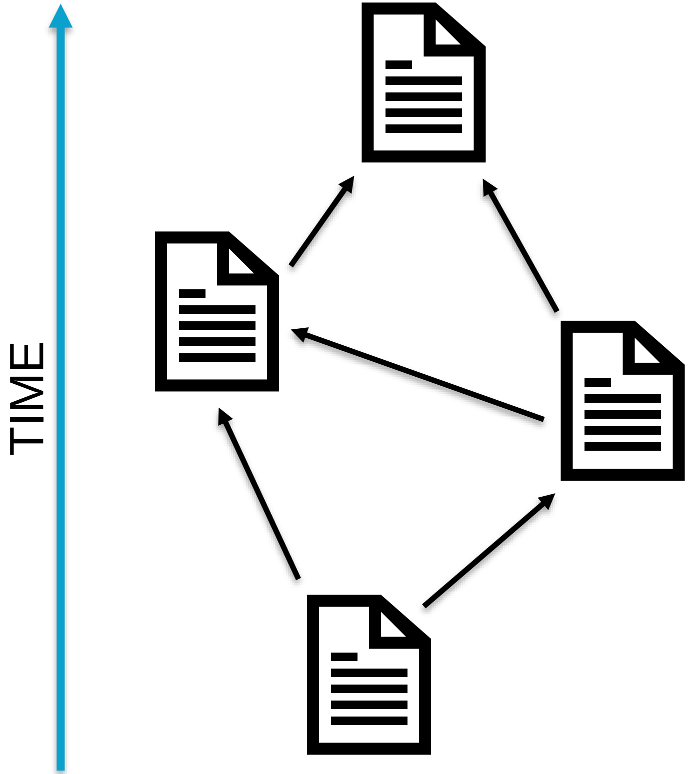
The Price model

Networks of Scientific Papers

The pattern of bibliographic references indicates the nature of the scientific research front.

(Science, 1965)

Derek J. de Solla Price

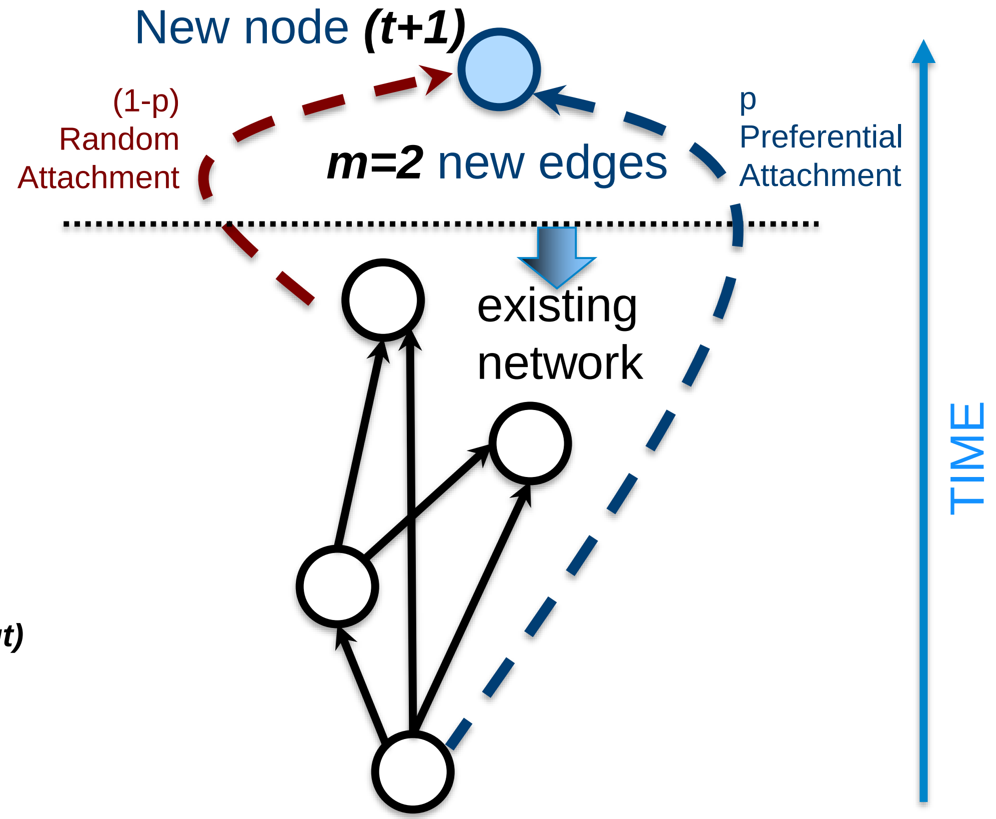


- Node = scientific publication
- Edge = from cited paper to citing paper
- Papers can only reference older papers, *arrow-of-time*
- Growing network model = Directed version of Barabási-Albert model

Price model

- Add new node ***(t+1)***
- Add ***m*** new edges to new node from existing node chosen using *either*
  - with probability ***p*** use Cumulative Advantage = Preferential Attachment Proportional to out-degree $k^{(out)}$

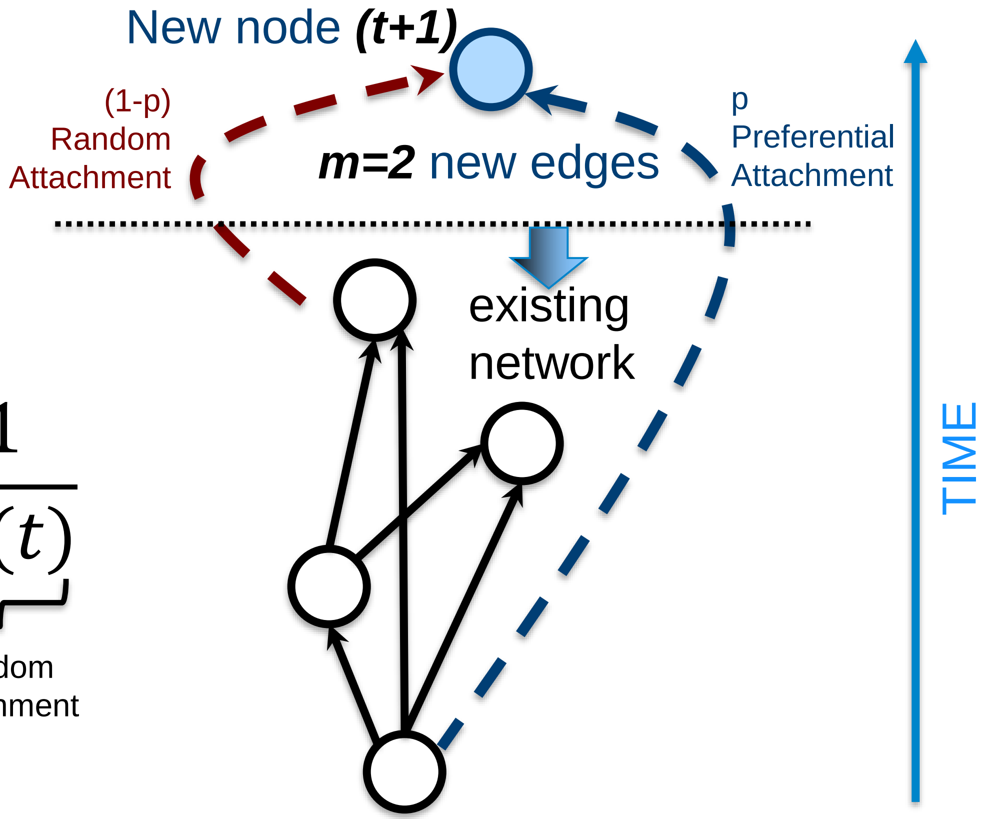
  or
  - with probability ***(1-p)*** use Random Attachment



Price model

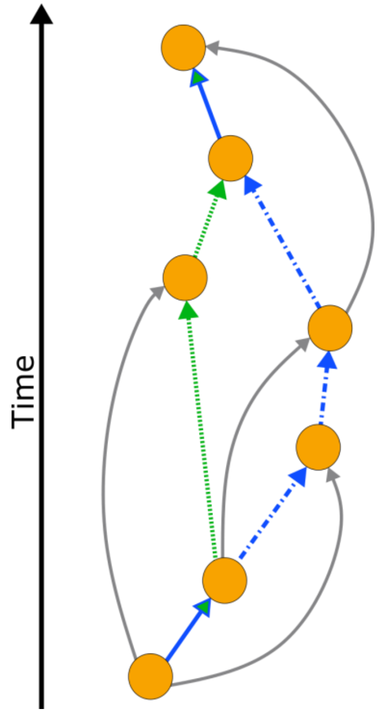
Choosing an existing node with out-degree $\boldsymbol{k}^{(out)}$ with probability Π

$$\Pi = p \underbrace{\frac{k^{(\text{out})}}{E(t)}}_{\substack{\text{cumulative advantage =}\\ \text{preferential attachment}}} + (1-p) \underbrace{\frac{1}{N(t)}}_{\substack{\text{random}\\ \text{attachment}}}$$

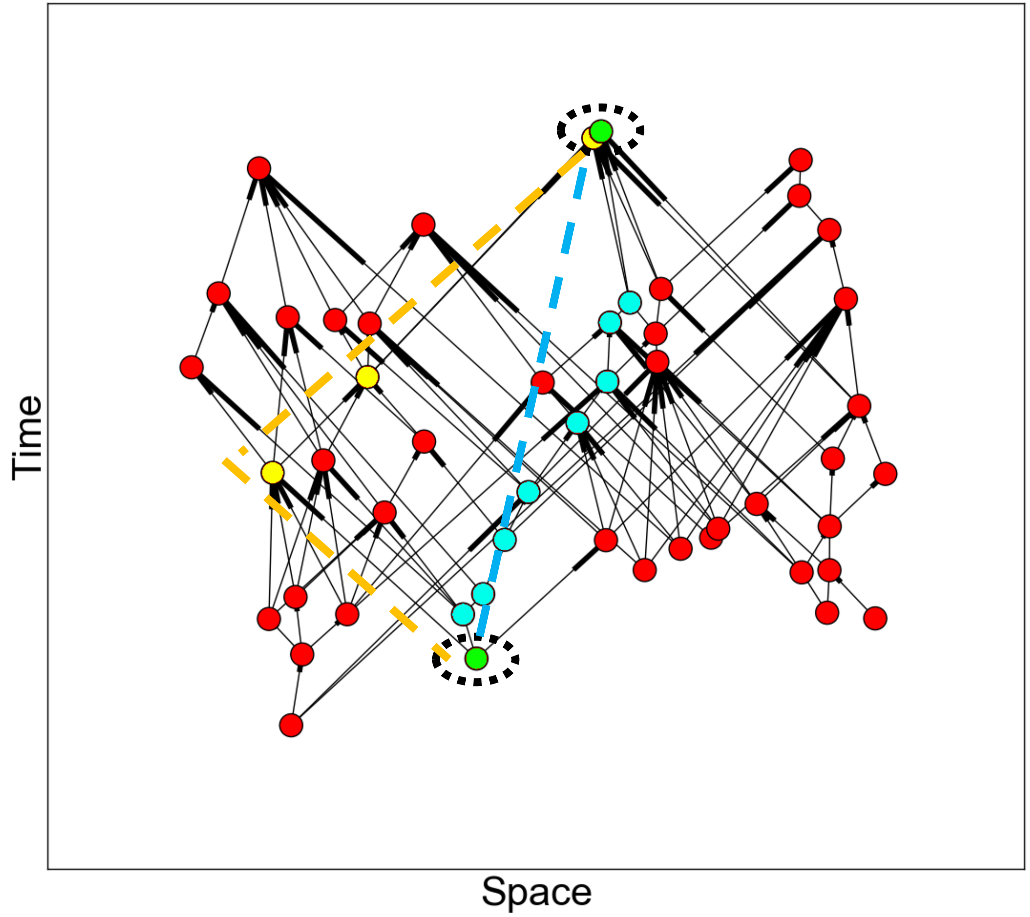


The Longest Path in Price Model

- The longest path is well-defined in a DAG
- Approximates geodesic in Minkowski random DAGs [Brightwell & Gregory 91]
- Similar to “Main Path” of bibliometrics?
- Longest path length ***L*** ~ ℓ greedy path length



Grey – SHORTEST =3
Green – GREEDY ℓ =4
Blue – LONGEST L= 5



Longest Path

Longest Path (9)

≈ Geodesic

Shortest Path (4)

≈ Edges of Light Cones

Minkowski PPP model

e.g. see Brightwell & Gregory 1991; Reid 2003; Evans and Clough 2016

The Greedy Path in Price Model

DEFINITION

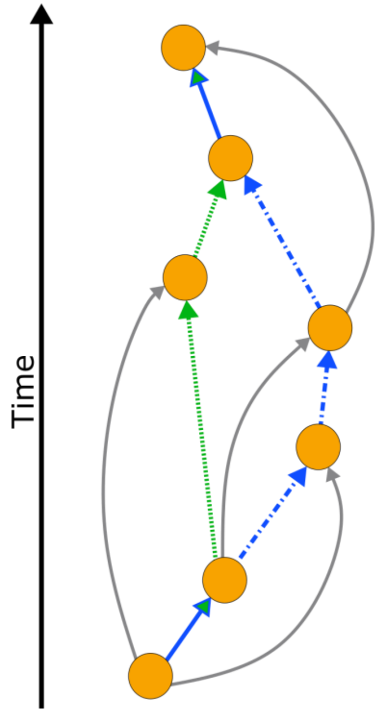
The next node on a Greedy Path
is the closest in time

Conjecture:

Longest path length ***L*** ~ ℓ greedy path length

$$\lim_{N \to \infty} \frac{\ell}{L} = c < 1$$

* Known for Minkowski Space PPP [Brightwell & Gregory 91]



Grey – SHORTEST =3
Green – GREEDY ℓ =4
Blue – LONGEST L= 5

The Master Equation for the length of the greedy path

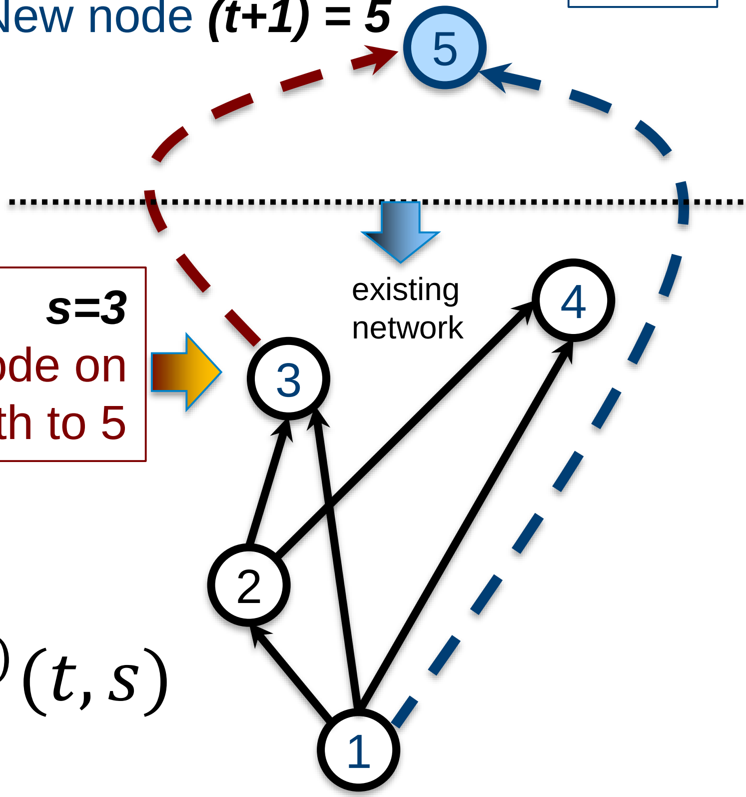
m=2 new edges

$\ell(\boldsymbol{t})$ = Length of greedy path
from source node ***t=1***
to node ***t***

$\Pi^{(max)}(t,s)$ = Probability ***s*** is largest time connected to ***t***

s=3
next node on greedy path to 5

New node ***(t+1) = 5***



Probability node at time ***(t+1)*** has length ℓ

$$P(\ell, t+1) = \sum_{s=1}^{t} P(\ell-1, s)\ \Pi^{(max)}(t,s)$$

The Master Equation for the length of the greedy path

$\ell(\boldsymbol{t})$ = Length of greedy path from source node ***t=1*** to node ***t***

Probability node at time ***(t+1)*** has length ℓ

$$P(\ell, t+1) = \sum_{s=1}^{t} P(\ell - 1, s)\ \Pi^{(max)}(t, s)$$

Probability that closest node chosen at time ***(t+1)*** is ***s***

$$\Pi^{(max)}(t, s) = \left(\Pi_{\leq}(t, s)\right)^m - \left(\Pi_{\leq}(t, s-1)\right)^m$$

cdf of attachment probability $\Pi_{\leq}(t, s) = \sum_{r=1}^{s} \Pi(t, r)$

The Longest Path in the Price Model

The Generating Function solution

$$G(z,t) = \sum_{z=0}^{\infty} z^{\ell} P(\ell,t)$$

$$G(z,t) = \prod_{s=1}^{t-1} \left[z + (1-z) \left(\frac{s-1+p}{s} \right)^{m} \right]$$

Simple linear form for attachment probability Π allows for exact solution at finite time within mean field approximation.

Generating function is a product of ***m*** Gamma function ratios

The Average Greedy Path Length

***m*=**
number of new edges at each t

$\bar{p} = (1-p)$ probability to connect to existing nodes using random attachment

$$\lim_{t\to\infty} \ell(t) = m\bar{p}\ln(t) - m\bar{p}\,\psi(m\bar{p}+1) + \sum_{n=2}^{m}\binom{m}{n}(-1)^{n-1}(\bar{p})^{n}\xi(n) + O(t^{-1})$$

digamma function

Riemann's zeta function

The Greedy Path Length Distribution

***m*=**
number of new edges at each t

$\overline{p} = (1-p)$ probability to connect to existing nodes using random attachment

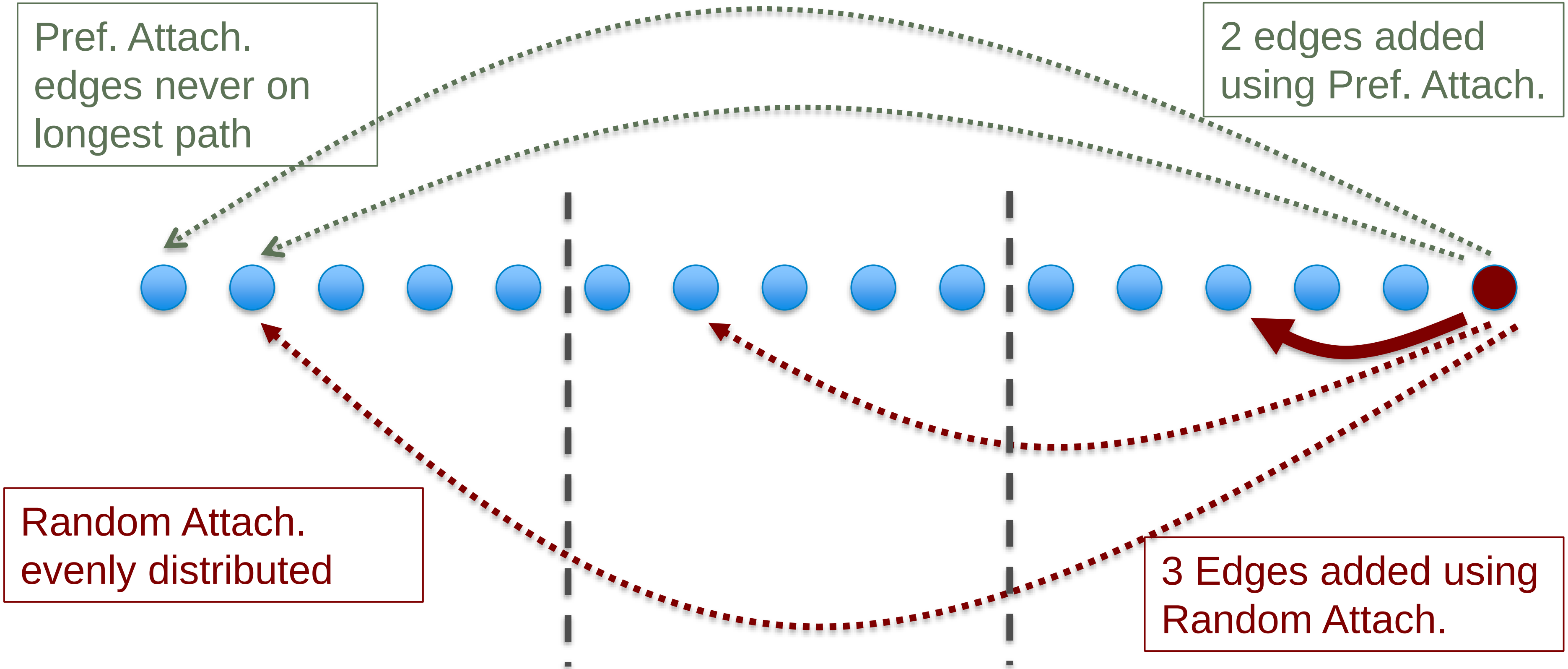
Average:

$$\lim_{t\to\infty} \ell(t) = m\,\bar{p}\,\ln(t) + O(1)$$

Distribution: **BINOMIAL** in long time limit

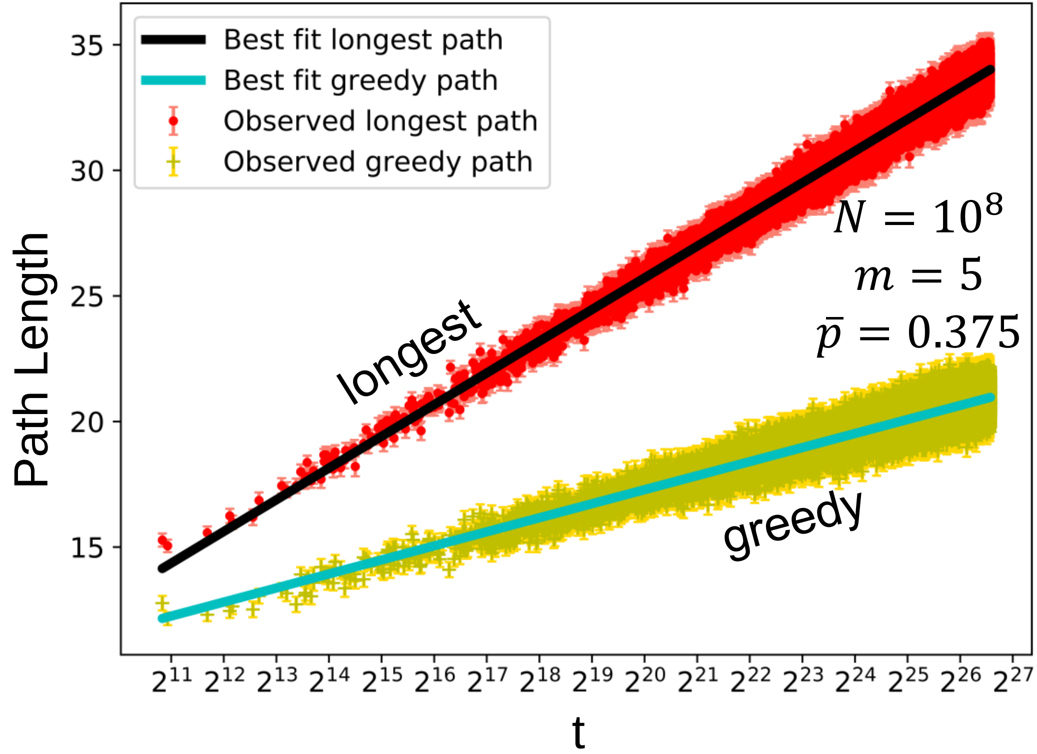
Random Attachment gives Longest Path

m=5, p=2/5



Numerical Example

$$\boxed{\ell(t) = a \ln(t) - b}$$



Average over 100 runs

Greedy path length scales very close to prediction:

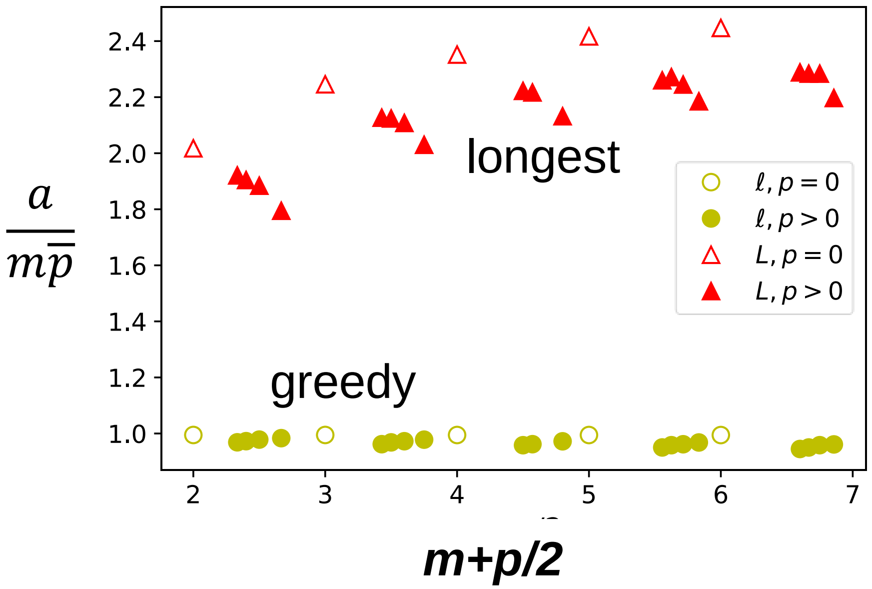
$$\frac{a_{\text{obs}}}{a_{\text{theo}}} = 0.96$$

For the longest path the slope is more than twice as large as for the greedy path:

$$\frac{a(L)}{a(\ell)} = 2.37$$

Numerical Results

Fit to $a \ln t + \mathrm{b}$

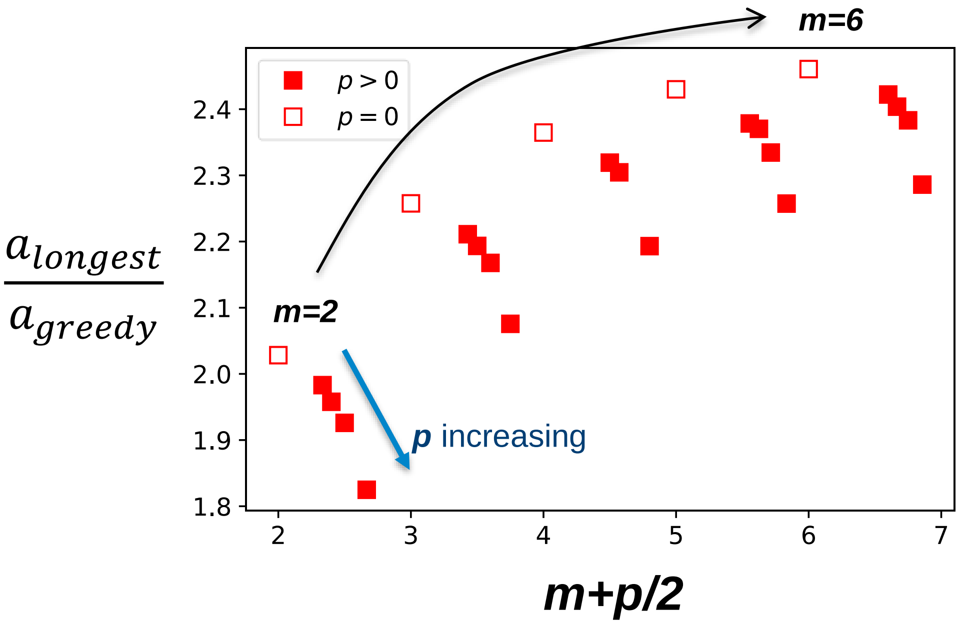


Numerical errors are smaller than symbols

Greedy path ℓ scales as $m\bar{p}$ ln(t)

Numerical Results

Fit to $a \ln t + b$



Numerical errors are smaller than symbols

Longest path $L \approx$ twice ℓ

 Tim EVANS, Imperial College

Summary

- An analytical solution for greedy path length in the Price model
- The length of longest and greedy paths in the Price model scale as ***log(N)*** in a network of ***N*** nodes
- The analytical and numerical results in excellent agreement.

Tim Evans,
Lucille Calmon,
Vaiva Vasiliauskaite

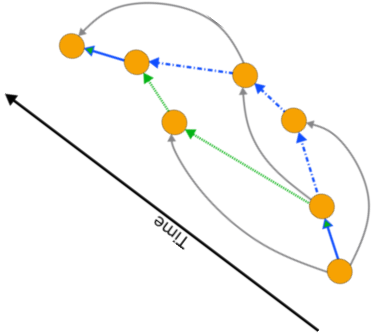
arXiv:1903.03667

Slides at

http://bit.ly/2oR62sI

DOI: 10.6084/m9.figshare.9933806

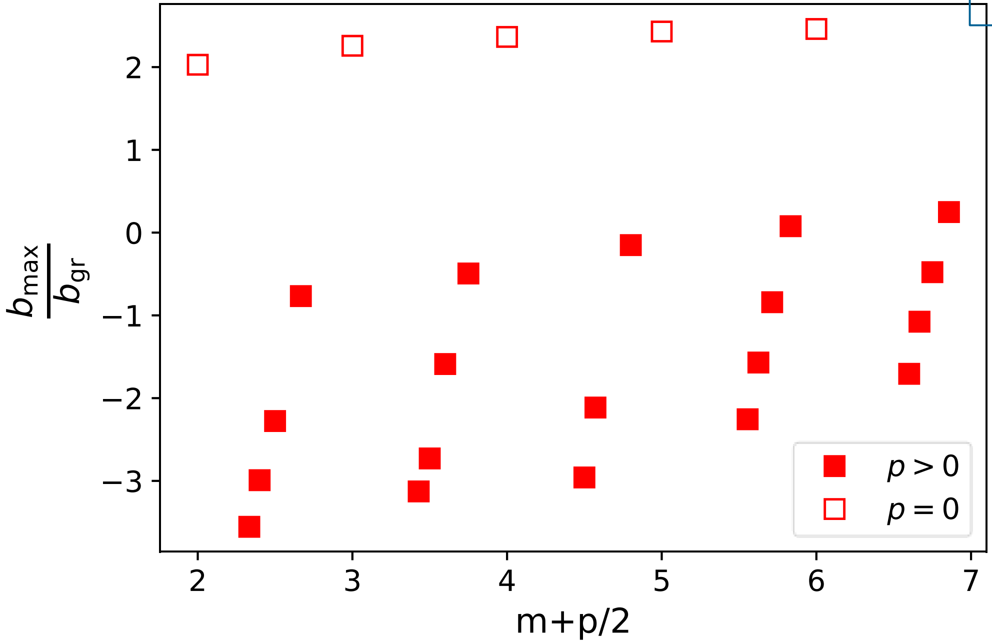
The arrow-of-time inherent in growing network models produces new distinctive features.



Extra Slides

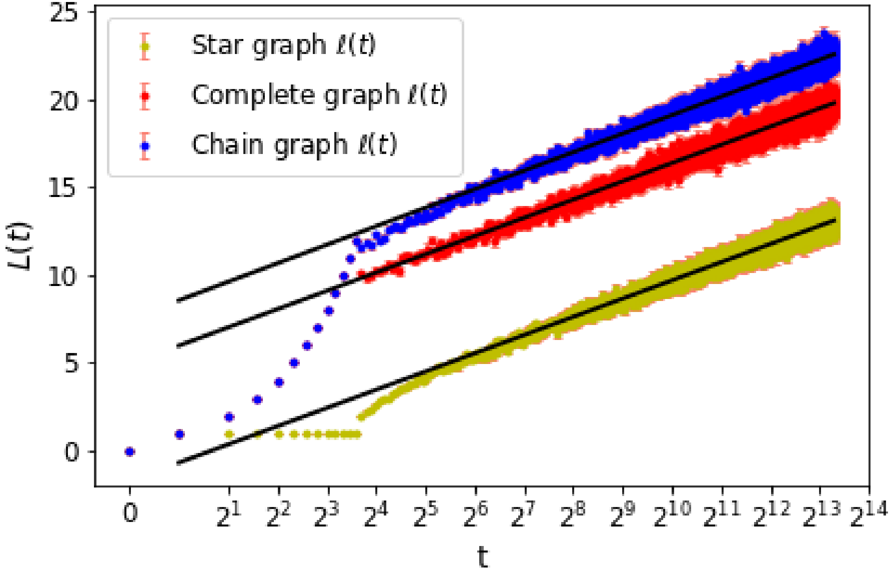
The second-to-leading order term

$$\ell(t) = a\ln(t) - b$$

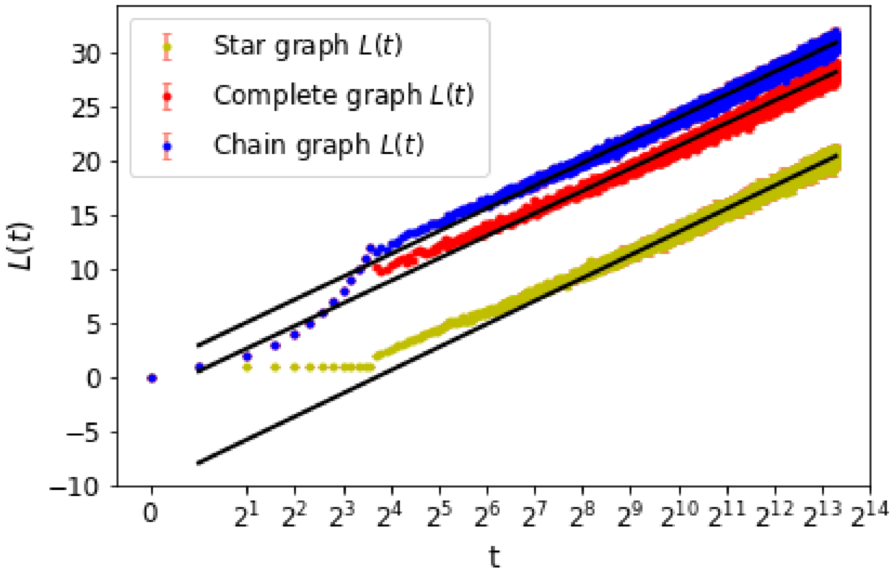


No clear scaling pattern

Initial graph effect



Greedy path ℓ scales as $m\bar{p}$



Longest path $L \approx$ twice ℓ

Just a shift in the constant

Bibliography

- L. Bombelli, Ph.D. thesis, Syracuse University, 1987.
- Brightwell, G. & Gregory, R. Structure of random discrete spacetime, *Phys. Rev. Lett.*, **1991**, *66*, 260-263
- Bruggemann, R.; Halfon, E.; Welzl, G.; Voigt, K. & Steinberg, C. Applying the Concept of Partially Ordered Sets on the Ranking of Near-Shore Sediments by a Battery of Tests J. Chem. Inf. Model., American Chemical Society (ACS), 2001, 41, 918-925.
- Bruggemann, R., M¨unzer, B. and Halfon, E., An algebraic/graphical tool to compare ecosystems with respect to their pollution – the German river “Elbe” as an example - I: Hasse-diagrams, Chemosphere, 28 (1994) 863–872.
- Clough, J. R.; Gollings, J.; Loach, T. V. & Evans, T. S., Transitive reduction of citation networks, J.Complex Networks, **2015**, 3, 189-203 [10.1093/comnet/cnu039 arXiv1310.8224]
- Clough, J.R. & Evans, T.S. What is the dimension of citation space? Physica A **2016,** *448*, 235-247 **[** arXiv:1408.1274 10.1016/j.physa.2015.12.053]
- Evans, T.S., Complex Networks, *Contemporary Physics,* **2004***, 45*, 455-474 [10.1080/00107510412331283531 arXiv:cond-mat/0405123]
- Evans, T.S., Clique Graphs and Overlapping Communities, J.Stat.Mech*,* **2010**, P12037 [10.1088/1742-5468/2010/12/P12037 arXiv:1009.0638]
- Evans, T.S. & Lambiotte, R., Line Graphs, Link Partitions and Overlapping Communities, Phys.Rev.E, **2009***, 80*, 016105 [10.1103/PhysRevE.80.016105 arXiv:0903.2181]
- Evans, T.S. & Lambiotte, R., Line Graphs of Weighted Networks for Overlapping Communities, Eur.Phys.J. B **2010***, 77*, 265–272 [10.1140/epjb/e2010-00261-8 arXiv:0912.4389]
- Expert, P.; Evans, T. S.; Blondel, V. D. & Lambiotte, R. Uncovering space-independent communities in spatial networks, *PNAS,* **2011***, 108*, 7663-7668 10.1073/pnas.1018962108 [10.1073/pnas.1018962108 arXiv:1012.3409]
- Garfield, E.; Sher, I. H. & Torpie, R. J., The use of citation data in writing the history of science, DTIC Document, 1964.
- Holme, P. & Saramäki, J., *Temporal Networks,* Physics Reports, 2012, 519, 97-125.
- Hummon, N. P. & Dereian, P., Connectivity in a citation network: The development of DNA theory, Social Networks, **1989**, 11, 39-63.
- J Myrheim. Statistical geometry, 1978. Technical report, CERN preprint TH-2538, 1978.
- D.A. Meyer. The Dimension of Causal Sets. PhD thesis, MIT, 1988.
- David A. Meyer. Dimension of causal sets, 2006.
- Price, D. J. de S. The scientific foundations of science policy, Nature **1965**, *206*, 233-238.
- Reid, D. D. Manifold dimension of a causal set: Tests in conformally flat spacetimes, Phys. Rev. D **2003***, 67*, 024034.