Dynamics on expanding spaces: modeling the emergence of novelties

Vittorio Loreto, Vito D. P. Servedio, Steven H. Strogatz and Francesca Tria

INFO-I501 | Introduction to Informatics | Paper Presentation | Kaicheng Yang

What is this paper about?

A review about modeling the emergence of innovations.

Models

<table>
<tr><td rowspan="4">Simon like models</td><td>Plain Simon’s</td></tr>
<tr><td>With time dependent sub-linear invention probability</td></tr>
<tr><td>With tokens aging</td></tr>
<tr><td>With stream aging</td></tr>
<tr><td colspan="2">Sample-space reducing model</td></tr>
<tr><td colspan="2">Hoppe Urn Model</td></tr>
<tr><td colspan="2">Urn model with triggering</td></tr>
</table>

Fundamental laws

Power law

Frequency distribution is power law

$$P(f) \propto f^{-\beta}, \beta > 1$$

Zipf's law

Frequency rank curve is power law

$$f(R) \propto R^{-\alpha}, \alpha = 1/(\beta - 1)$$

Heap's law

Size of distinct elements grows sub-linearly with total number of elements

$$D \propto \begin{cases} N^{\gamma}, & \beta < 2 \\ N, & \beta > 2 \end{cases}, \gamma = \beta - 1$$

Real data:

$$\beta < 2$$

$$\alpha > 1$$

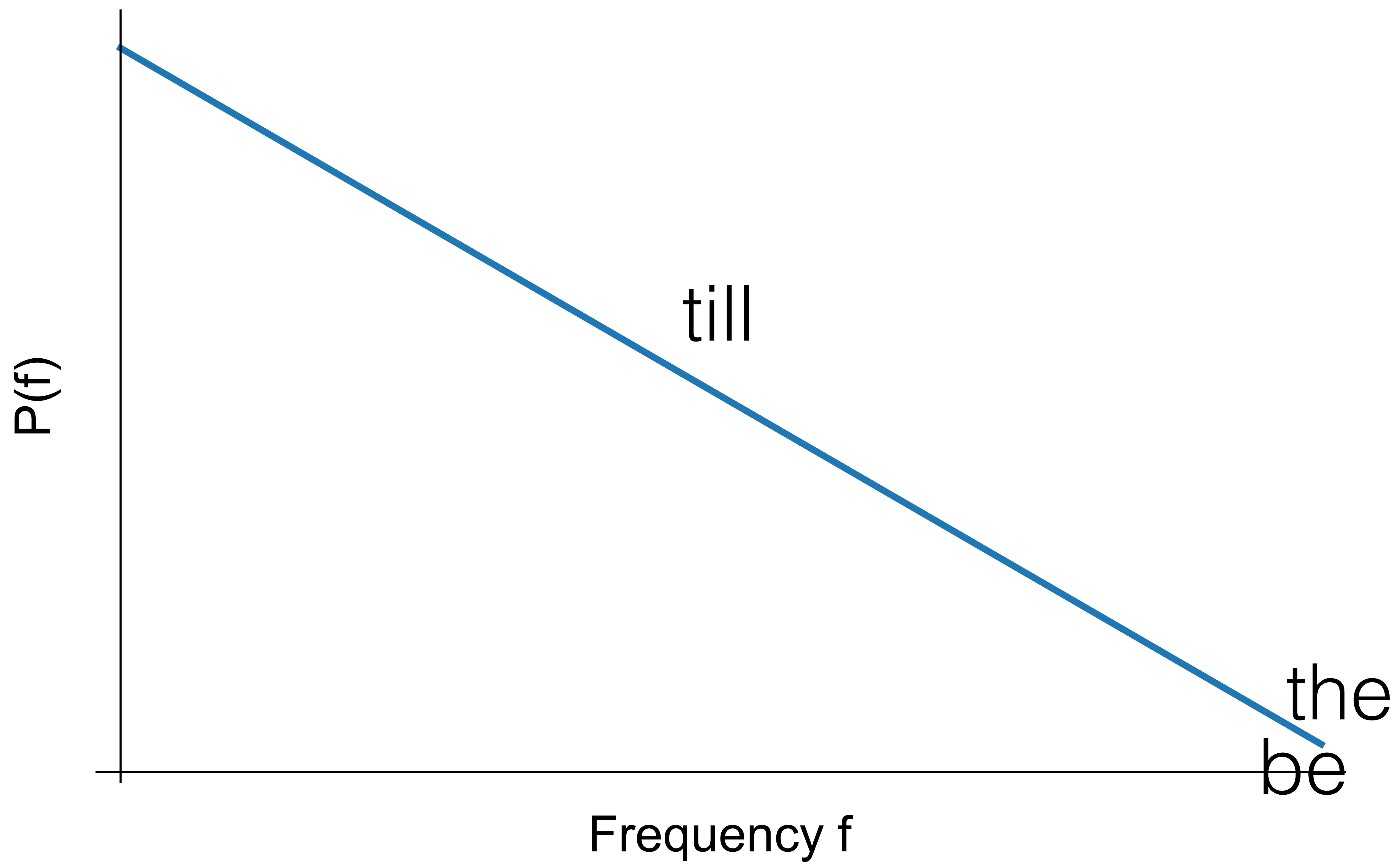
$$\gamma < 1$$

Words list

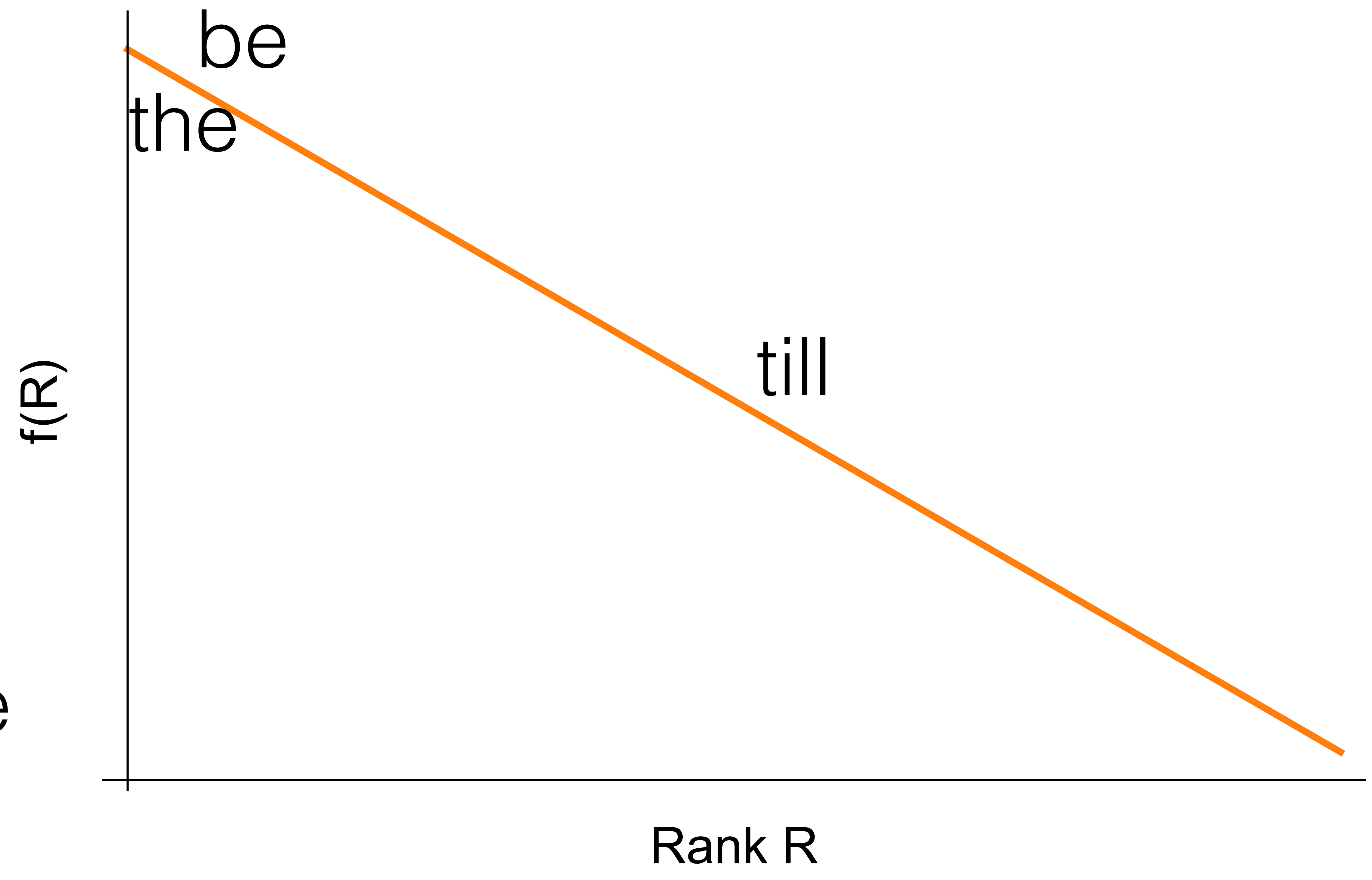
Word	Frequency	Rank
the	22038615	1
be	12545825	2
and	10741073	3
…	…	…
till	5079	5000
...	...	...

Data from: https://www.wordfrequency.info/

Power law and Zipf's law

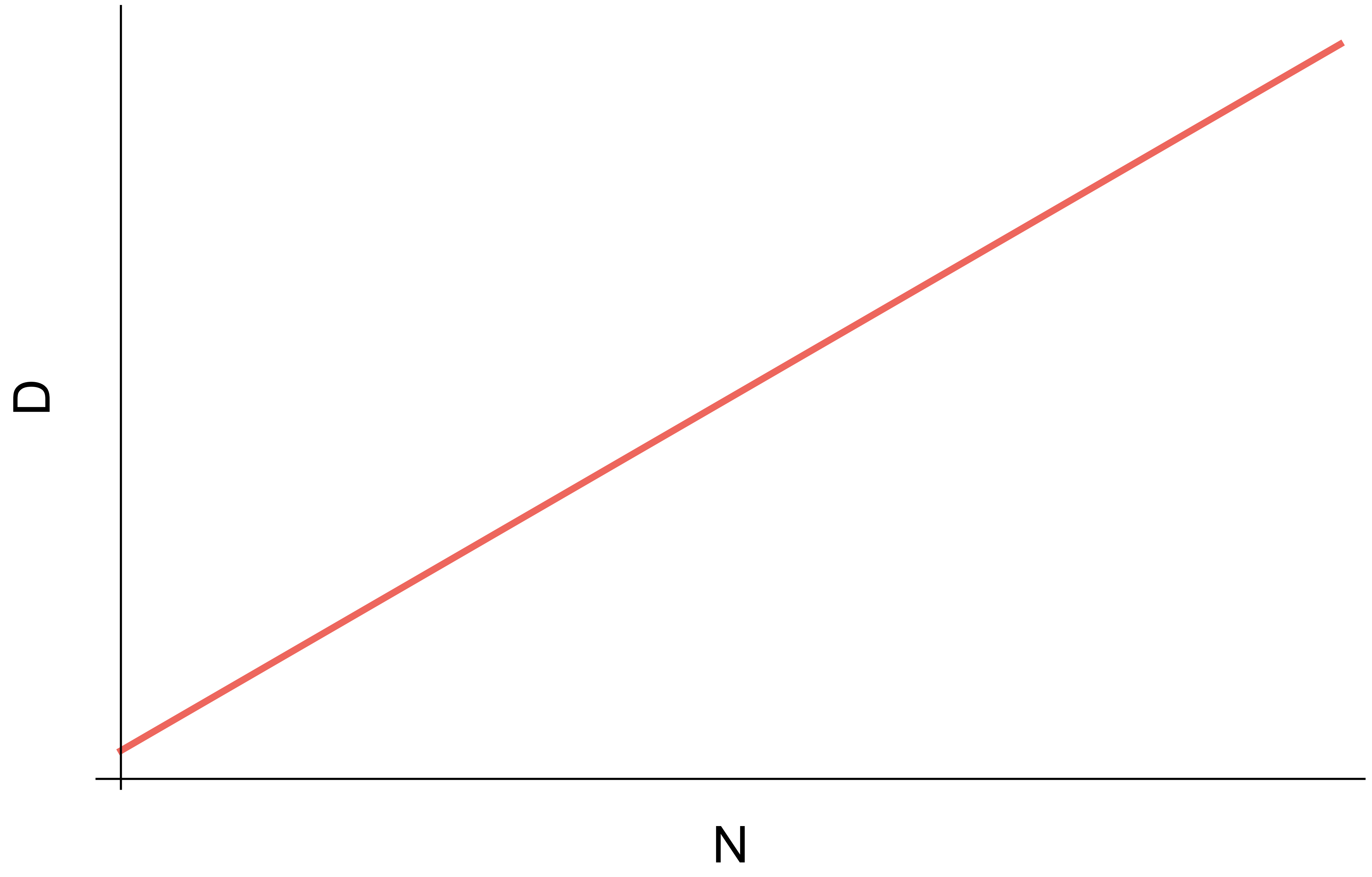


Frequency distribution



Frequency rank curve

Heap's law



D: number of distinct elements
N: total number of elements

Fundamental laws

Power law

Frequency distribution is power law

$$P(f) \propto f^{-\beta}, \beta > 1$$

Zipf's law

Frequency rank curve is power law

$$f(R) \propto R^{-\alpha}, \alpha = 1/(\beta - 1)$$

Heap's law

Size of distinct elements grows sub-linearly with total number of elements

$$D \propto \begin{cases} N^{\gamma}, & \beta < 2 \\ N, & \beta > 2 \end{cases}, \gamma = \beta - 1$$

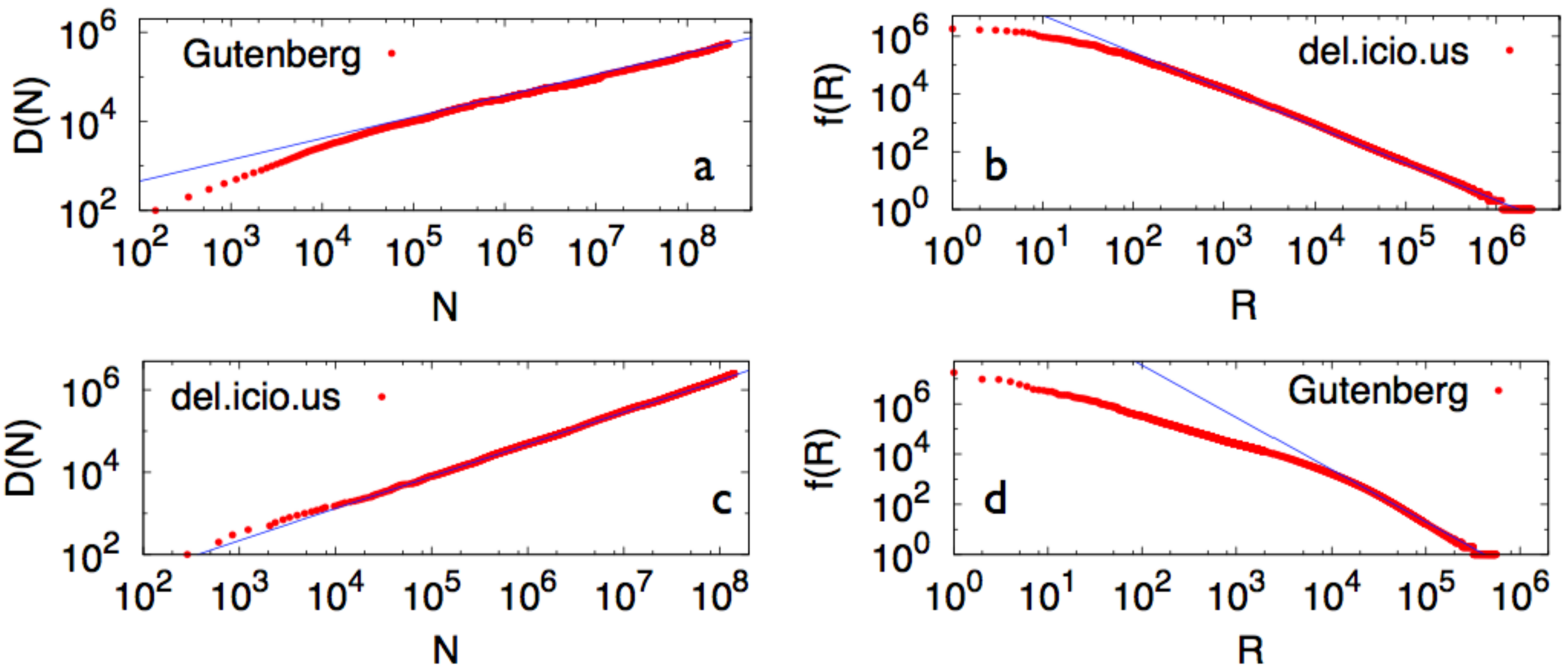
Real data:

$$\beta < 2$$

$$\alpha > 1$$

$$\gamma < 1$$

Fundamental laws



Heap's law: $D(N) \propto N^{\gamma}$

Zipf's law: $f(R) \propto R^{\alpha}$

Plain Simon's model

1. A stream of tokens
2. Each time a new token in added to the stream with probability p
3. With probability (1-p) a randomly chosen token in the stream is selected to be copied

rich-gets-richer → Power law distribution

Plain Simon's model

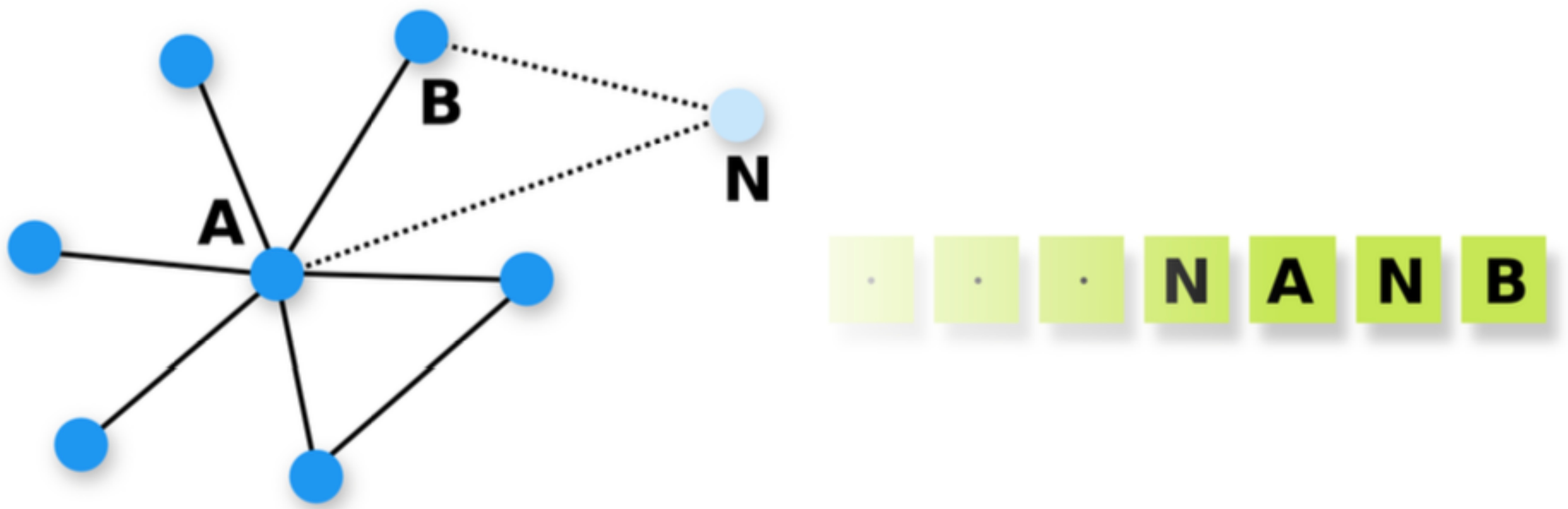
Master equation: $$N_{k,t+1} = N_{k,t} + (k-1)(1-p)\frac{N_{k-1},t}{t} - k(1-p)\frac{N_{k,t}}{t} + p\delta_{k,1}$$

Continuous limit: $$\frac{\partial N_k}{\partial t} = -(1-p)\frac{1}{t}\frac{\partial (kN_k)}{\partial k} \qquad N_{k,t} = tq_k$$

Solution: $$q_k \propto k^{-1-\frac{1}{1-p}} \qquad \beta = 1 + \frac{1}{1-p} \qquad p \in (0,1)$$

$\beta \in (2,\infty)$ Linear growth, not sub linear

Equivalence with BA model



Models	BA	Simon's
Equivalence quantity	Degree distribution	Frequency distribution
Settings	m	p=1/2

Simon's model with sub-linear p

To counter Plain Simon's model's linear growth, the invention probability should decrease sub-linearly:

$$p(t) = p_0 t^{\gamma - 1}, 0 < \gamma < 1$$

Sub-linear growth: $D(t) \propto t^{\gamma}$

Zipf's law: $f(R) \propto R^{1/\gamma}$

Simon’s model with memory

Old songs are not likely to be chosen, so tokens should age.
Different aging mechanisms were introduced by Dorogovtsev-Mendes and Cattuto-Loreto-Pietronero.

Still, linear growth

The Dorogovtsev-Mendes model:

1. Introduced as an improvement of BA model to create scale-free networks.
2. Should lead to a similar result for modeling emergence of novelties due to the equivalence.

Sample-space reducing model

“Space of possibilities often locally reduces when the process goes on: for instance, when composing a sentence the first word is almost free, while the subsequent ones are more and more constrained.”

The model:

1. N-faced dice => j
2. j-faced dice => i
3. i-faced dice => 1
4. go to rule 1
5. each step go to rule 1 with λ

Power law without rich-get-richer mechanism

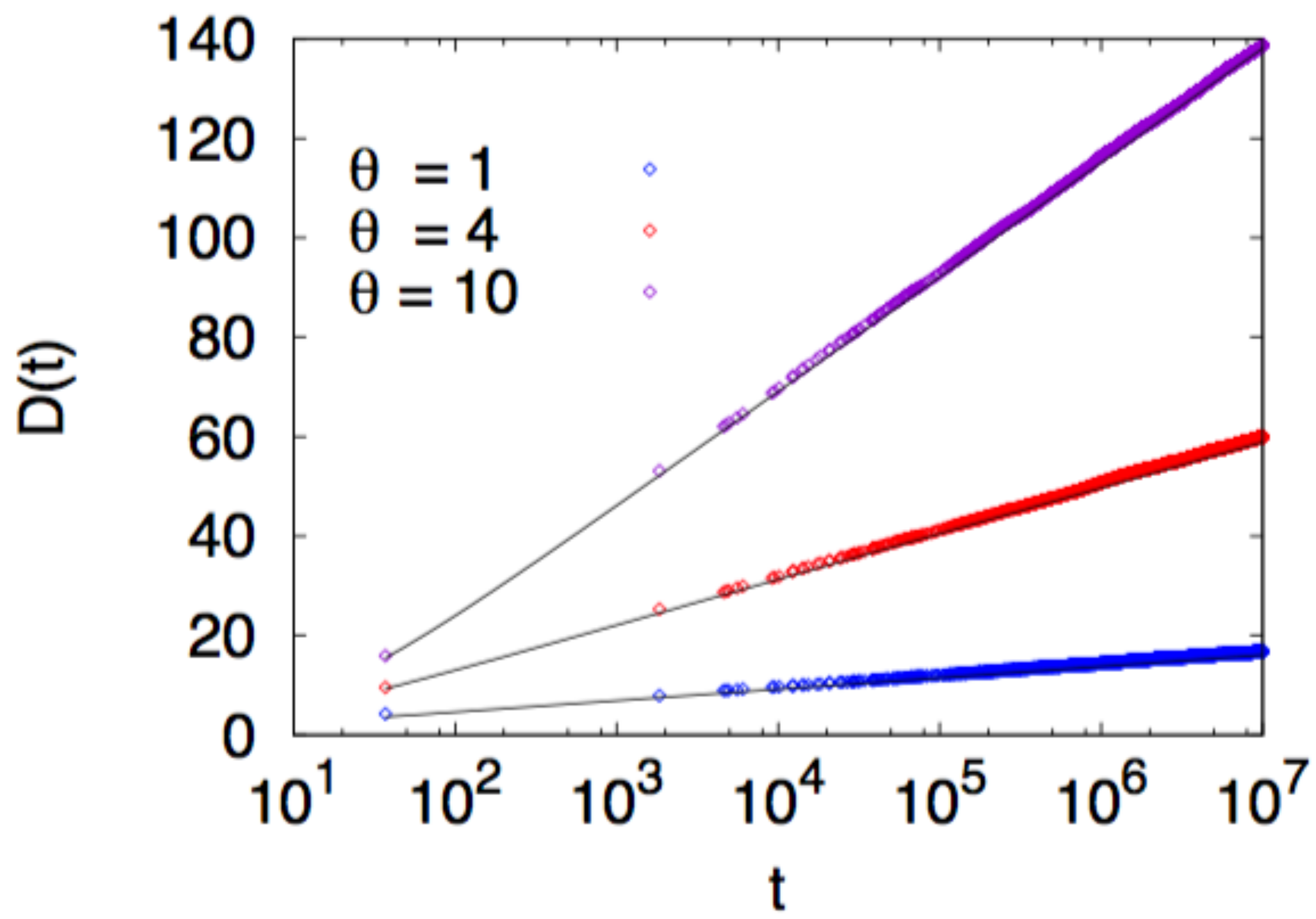
$$f(R) \propto R^{-\lambda}$$

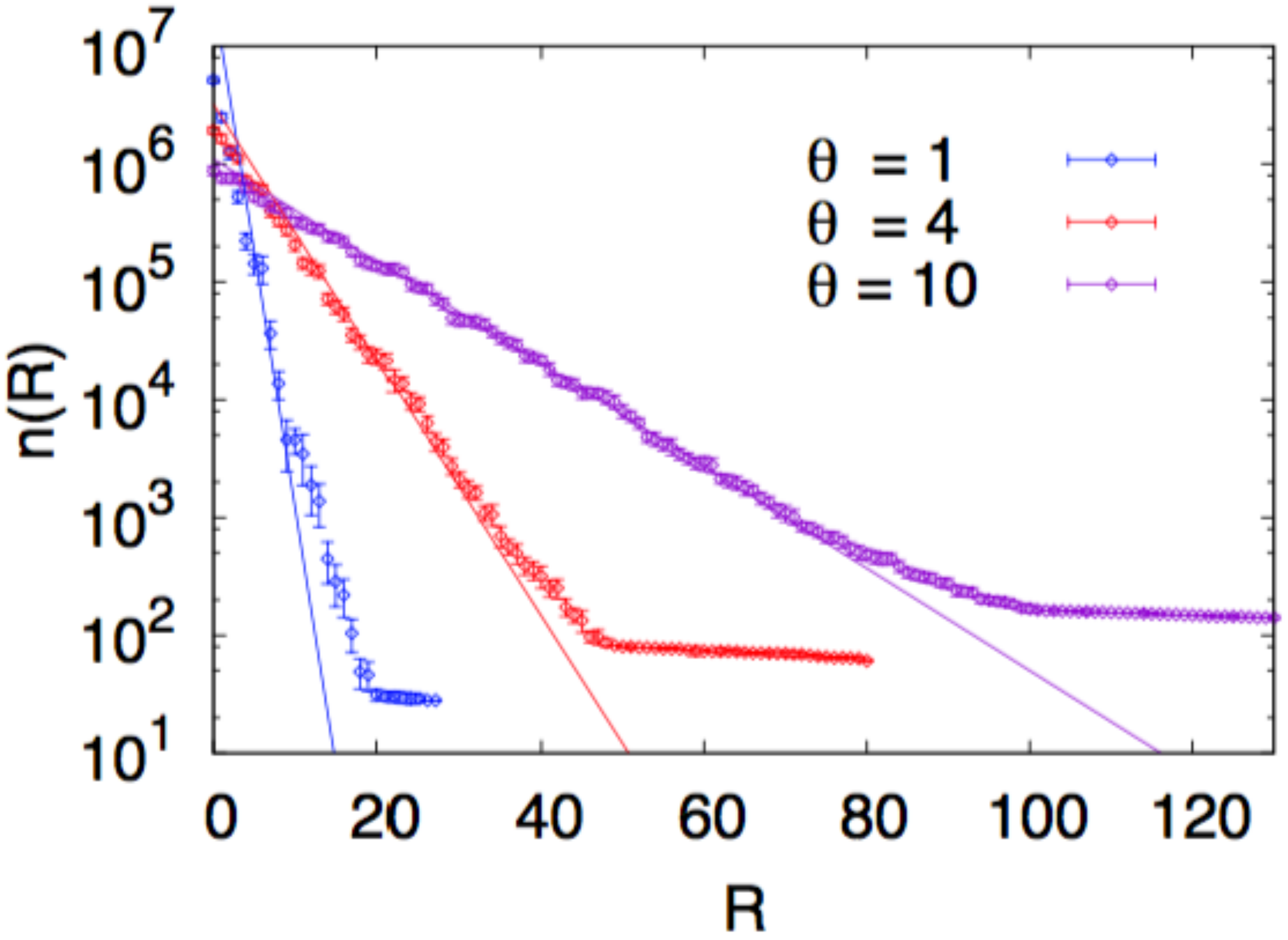
$$\lambda \in (0, 1)$$

Hoppe urn model

1. Two kinds of balls: black ones with mass θ colored with mass 1
2. Start with one black ball in the urn
3. Randomly choose balls from the urn, if it’s black return it and add a ball with new color; if it’s colored, return it with a copy of it
4. Repeat 3

Hoppe urn model

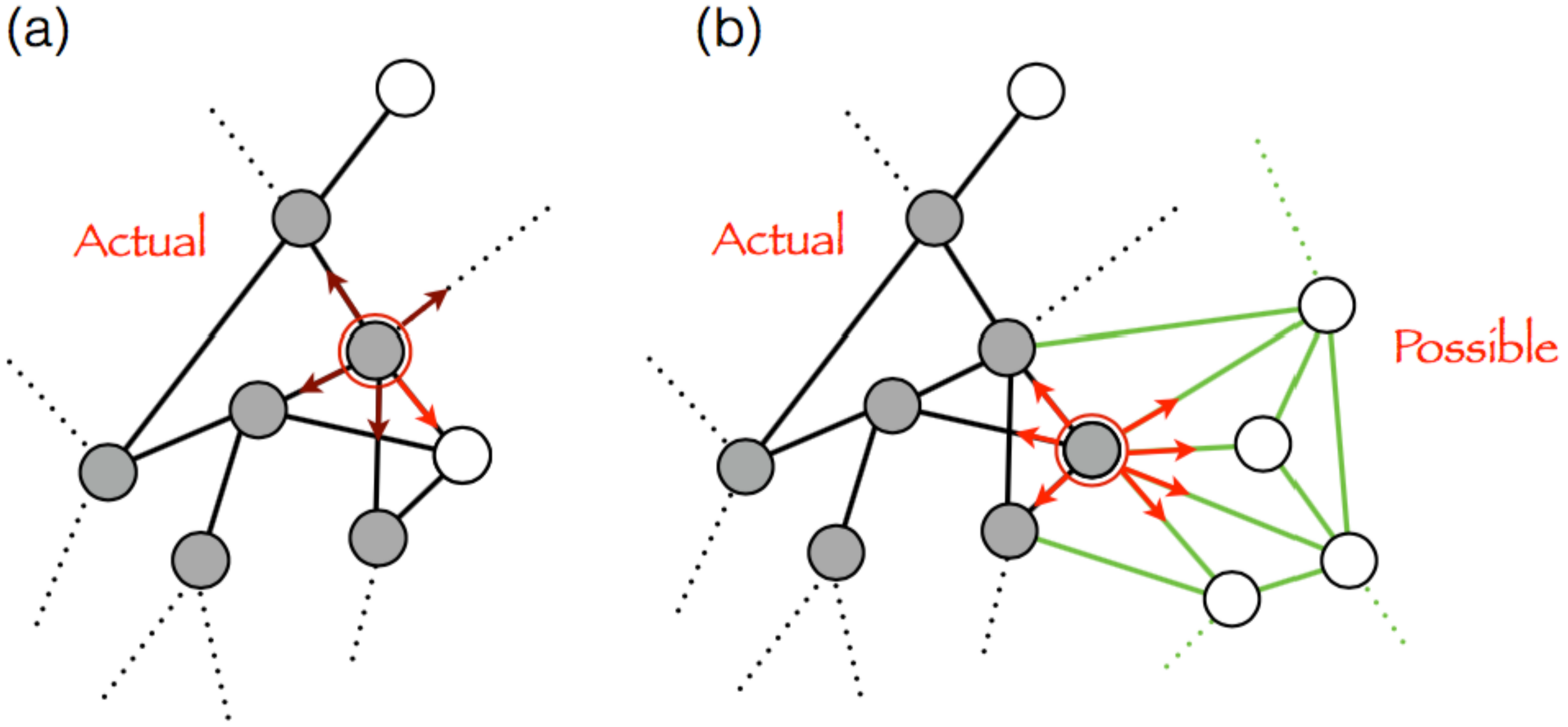




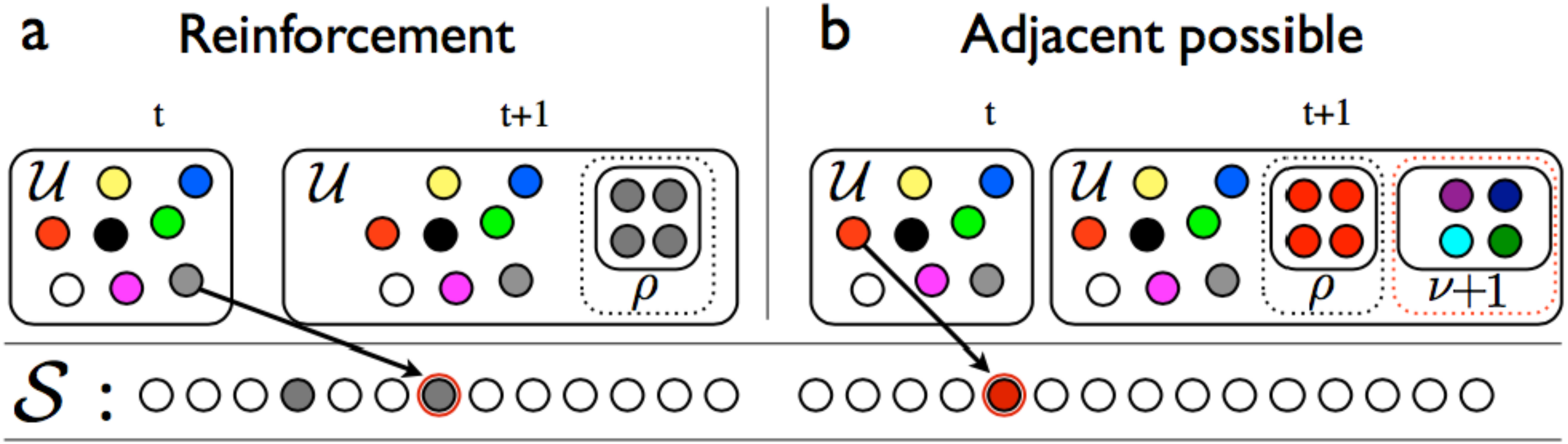
Heap's law: $D(t) = \theta \ln\left(1 + \frac{t}{\theta}\right)$

Zipf's law: $n(R) \approx \frac{t}{\theta} \exp\left(-\frac{R-1}{\theta}\right)$

Urn model with triggering



Urn model with triggering



1. an element is randomly extracted from *U* with uniform probability and added to *S*
2. the extracted element is put back into *U* together with ρ copies of it
3. if the extracted element has never been used before in *S* (it is a new element in this respect), then ν + 1 different brand new distinct elements are added to *U*.

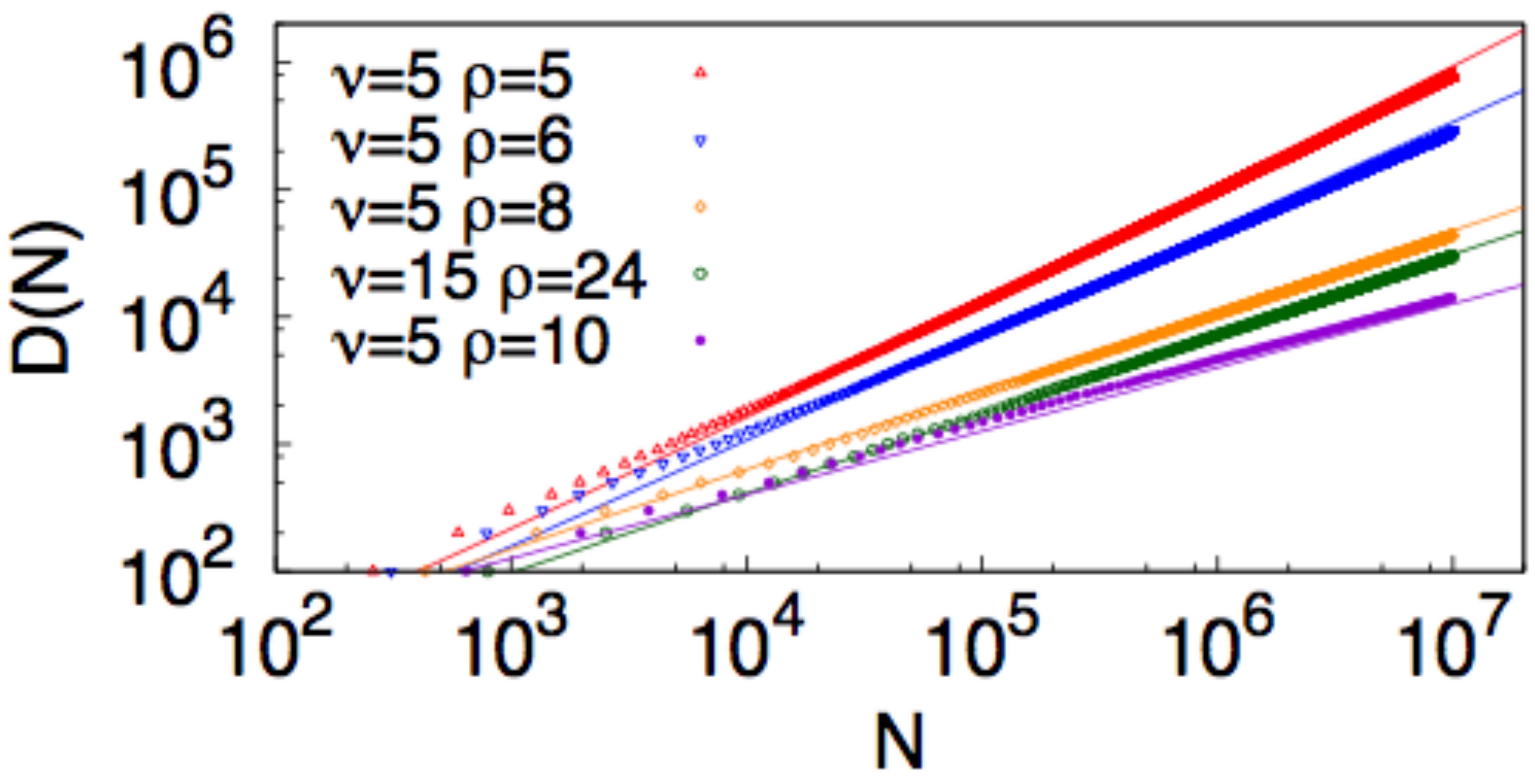
Urn model with triggering

$$D \propto \begin{cases} (\rho - v)^{\frac{v}{\rho}} t^{\frac{v}{\rho}}, & \rho > v \\ \frac{v-\rho}{a} t, & \rho < v \end{cases}$$

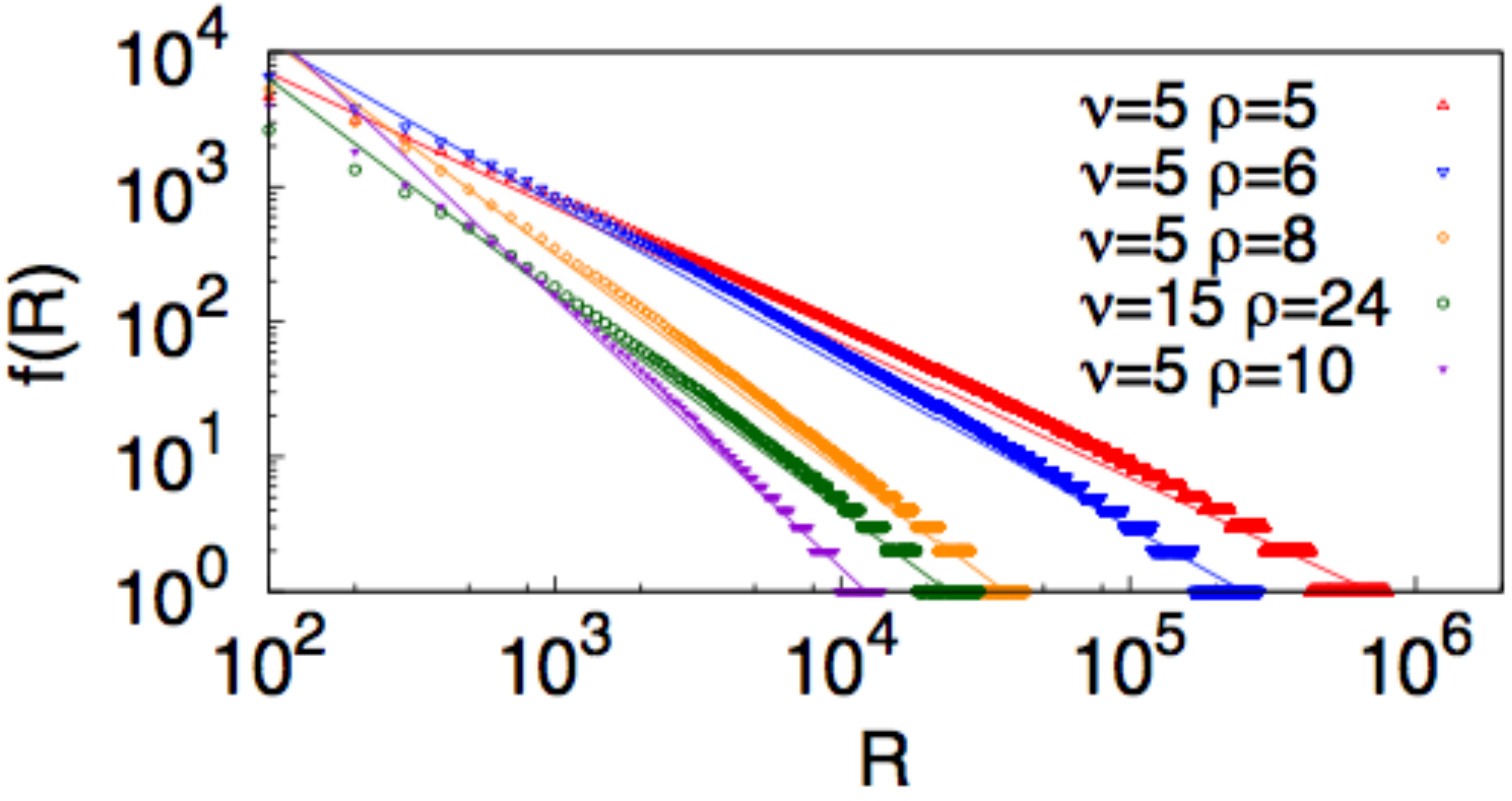
$$f(R) \propto R^{-\frac{\rho}{v}}$$

Recover the results to Plain Simon's model when $\rho < v$

Urn model with triggering



Heap's law



Zipf's law

Discussions

1. Are you convinced that this is the right approach?
2. Why the paper reviews the models that don’t work well?
3. How can these models help us to make more innovations?

Thanks!