The fate of leaders in growing networks

Jean-Marc Luck

Institut de Physique Théorique (CEA & CNRS), Saclay, France

Collaboration with Claude Godrèche and Hélène Grandclaude

- J. Stat. Mech. P11006 (2008)
- J. Stat. Phys. **137**, 1117 (2009)
- J. Stat. Mech. P02001 (2010)
- J. Stat. Mech. P07031 (2010)

Plan of this talk

Background

- Complex networks
- Why stochastic growth models?
- Our favorite: the Bianconi-Barabási model
- Why study leaders and lead changes

Results

- $T = \infty$ *(Barabási-Albert model)*
- $T = 0$ *(Record-driven growth process)*
- T finite *(Generic Bianconi-Barabási model)*

Complex networks

- Have become a fashion in the 2000s *(like fractals in the 1980s)*

Natural	*Leaves, trees, rivers, lungs, blood vessels*
Man-made	*Roads, railways, airlines, Internet, WWW*
Artificial	*Chemical, biological, social, etc.*

- Provide a convenient tool to deal with huge sets of data

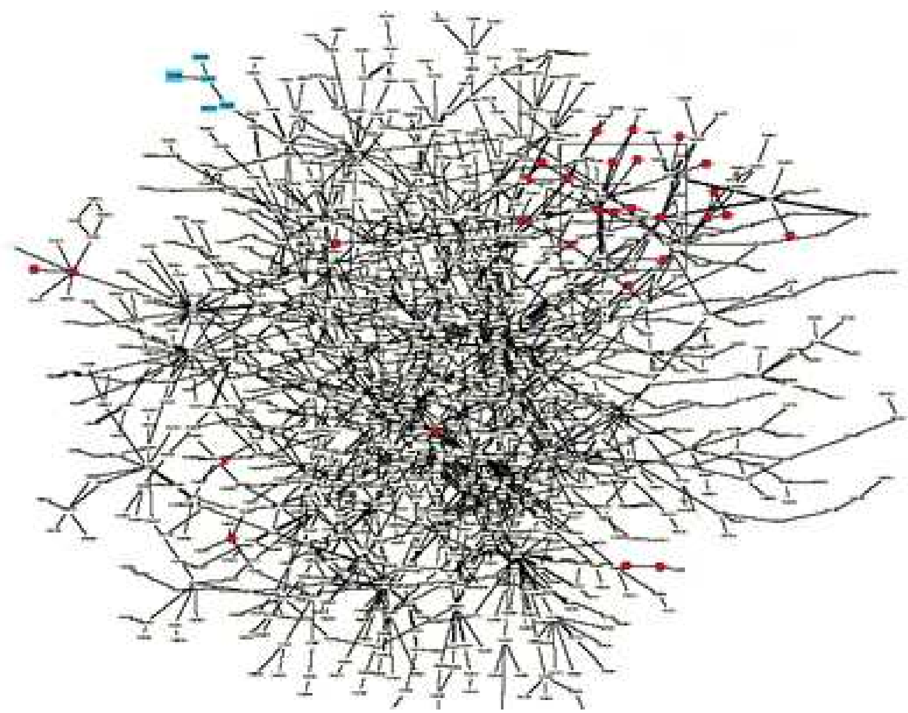
  Nodes = agents, links = relationships

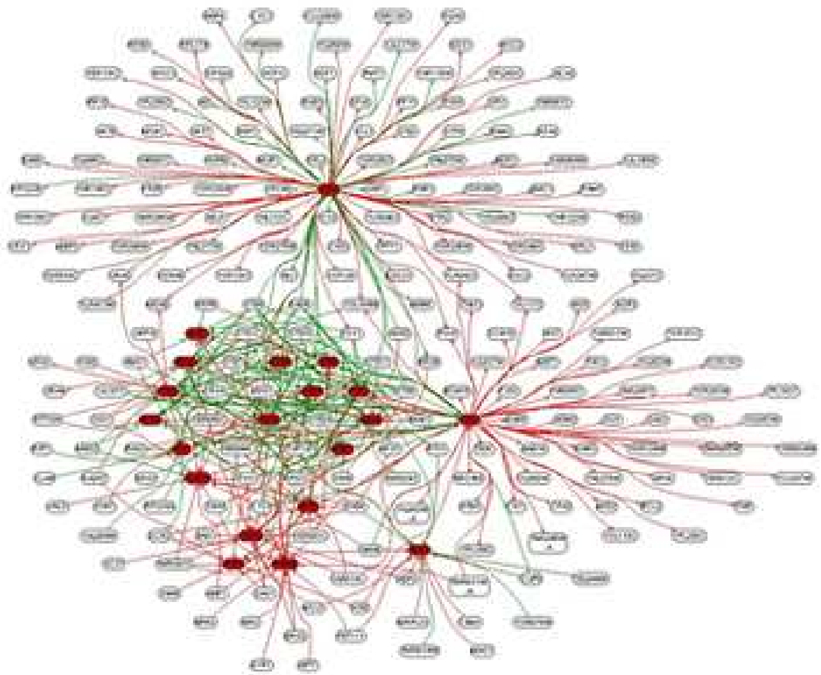
- Growing *(nonequilibrium)* rather than static *(equilibrium)* structures

  Many examples of networks do grow with time *(airline, Internet)*

  Most successful models of complex networks are *stochastic growth models*

Two random examples of artificial complex networks (biology)





Why stochastic growth models?

Most complex networks exhibit two key features

- *Scalefreeness*

  Broad distribution of node degree k, falling off as power law: $P(k) \sim k^{-\gamma}$

  Often $2 < \gamma < 3$, so that degree fluctuations diverge: $\operatorname{var} k \sim n^{(3-\gamma)/(\gamma-1)}$

- *Small-world effect*

  Slow growth of diameter of network: $\ell \sim \ln n$

  Effective dimensionality is infinite

- Static models

  generalizing Erdös-Rényi random graph model

  ... do not explain above key features in a natural way

- Growth models with preferential attachment

  Random growing tree

  Node n joins the network at time n

  Attaches to *single* earlier node $(m = 1, \ldots, n-1)$

  Attachment probability $\Pi_{n \to m}$ defines the model

  ... have proved far more successful

  ... have their own interest from the viewpoint of Statistical Physics

  ... may e.g. exhibit a non-equilibrium condensation transition

Our favorite: the Bianconi-Barabási model

(Bianconi & Barabási, 2001)

Attachment probability $\Pi_{n \to m} \sim k_m(n-1)\, q_m$

- Two ingredients

  Degree (i.e., popularity) $k_m(n-1)$ *Rich-get-richer*

  Fitness (i.e., quality) q_m *Fit-get-richer*

- Fitnesses assumed to be activated: $q_m = \exp(-\varepsilon_m/T)$

  Temperature T is measure of ruggedness of fitness landscape

  Activation energies $\varepsilon_m > 0$ quenched random variables with density $\rho(\varepsilon)$

  Condensation transition at critical temperature T_c *for some distributions* $\rho(\varepsilon)$

Why study leaders and lead changes

Leader at time n is node I whose degree is maximal: $k_I(n) = k_{\max}(n)$

Interest in *history of leaders and lead changes* ?

- Degree of node = *activity of airport, popularity of website, wealth of firm, ...*

  Problem is interesting *per se* (Krapivsky & Redner, 2002)

- Condensed phase ($T < T_c$ in BB model)

  Provides probe for *non-equilibrium dynamics of condensate*

Two classes of quantities

- *Local:* Degree $k_{\max}(n)$ and index $I(n)$ of leader *at* time n
- *Non-local:* Numbers $\mathcal{L}(n)$ of leads and $D(n)$ of distinct leaders *up to* time n

Results I: $T = \infty$

Barabási-Albert model (Barabási & Albert, 1999)

Attachment probability $\Pi_{n \to m} = \dfrac{k_m(n-1)}{Z(n-1)}$

- Minimal model with *linear rich-get-richer effect*
- Its success launched the fashion for scalefree networks... among physicists
- Analytical investigations possible
- Numerical simulations too *(redirection algorithm)*

Worth considering two initial states

(A) *Isolated node:* Degree $k_1(1) = 0$, hence $Z(n) = 2L(n) = 2n - 2$

(B) *Rooted node:* Degree $k_1(1) = 1$, hence $Z(n) = 2L(n) = 2n - 1$

How exact results yield heuristic predictions

- Degree distribution: $f_k = \dfrac{4}{k(k+1)(k+2)}$

  Scalefree with exponent $\gamma = 3$

  Extreme-value statistics argument yields $k_{\max}(n) \sim n^{1/2}$

- Mean degree of node i at time n :

  $$\left\langle k_i^{(\mathrm{A})}(n) \right\rangle = \frac{\Gamma(n-1/2)\,\Gamma(i-1)}{\Gamma(i-1/2)\,\Gamma(n-1)}, \quad \left\langle k_i^{(\mathrm{B})}(n) \right\rangle = \frac{\Gamma(n)\,\Gamma(i-1/2)}{\Gamma(i)\,\Gamma(n-1/2)}$$

  Scales as $\langle k_i(n) \rangle \approx \left(\dfrac{n}{i}\right)^{1/2}$

  Confirms $k_{\max}(n) \sim n^{1/2}$ and suggests microscopic $I(n) \sim 1$

  Fluctuations? Memory of initial state?

Rule: Initial state matters

Many quantities (both local and global)

depend on initial condition and have non-trivial fluctuations

Examples:

Reduced degree of leader $Y(n) = k_{\max}(n)/n^{1/2}$

Index $I(n)$ of leader

Number $D(n)$ of distinct leaders

Lead survival probability $S(n)$

Quantity	$\langle Y \rangle$	$\langle I \rangle$	$\langle D \rangle$	S
Case A	2.00	3.40	2.22	0.279
Case B	2.16	2.67	1.94	0.389

Summary of I

Typical history for $T = \infty$ (Barabási-Albert model)

- *Finitely many distinct leaders*
- *Among the few oldest nodes*
- *Finite probability of having a single leader*
- *Otherwise lead changes occur at logarithmic pace:* $\mathcal{L}(n) \sim \ln n$
- *All this depends on initial state*

Results II: $T = 0$

Record-driven growth process (Godrèche & L, 2008)

- Reminder: Fitnesses are activated: $q_m = \exp(-\varepsilon_m/T)$

  $T \to 0$ limit: $\varepsilon_m < \varepsilon_l$ implies $\Pi_{n\to m}/\Pi_{n\to l} \to \infty$

- Node N whose energy ε_N is smallest at time n : $\varepsilon_N = \min(\varepsilon_1, \ldots, \varepsilon_n)$

  Defined as the current record node at time n

  Attracts all new connections... until outdone by next record

- There is an infinite sequence of record times (i.e., record nodes)

  $\{N_1,\ N_2,\ N_3, \ldots\}$

- Only record nodes grow

- *Leaders* (largest degree) are among *records* (best fitness)

The universal (but not so well-known) statistics of records

Time N is a record if $\varepsilon_N < \min(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_{N-1})$

- This occurs with prob. $p_N = \dfrac{1}{N}$ *irrespective of past, future, and* $\rho(\varepsilon)$
- Number of records up to time n grows as $\ln n$

Sequence of record times $\{N_1 = 1,\ N_2,\ N_3, \ldots\}$

- Recursive construction

  If N is a record, next record is M with prob. $p_{N,M} = \dfrac{N}{M(M-1)}$

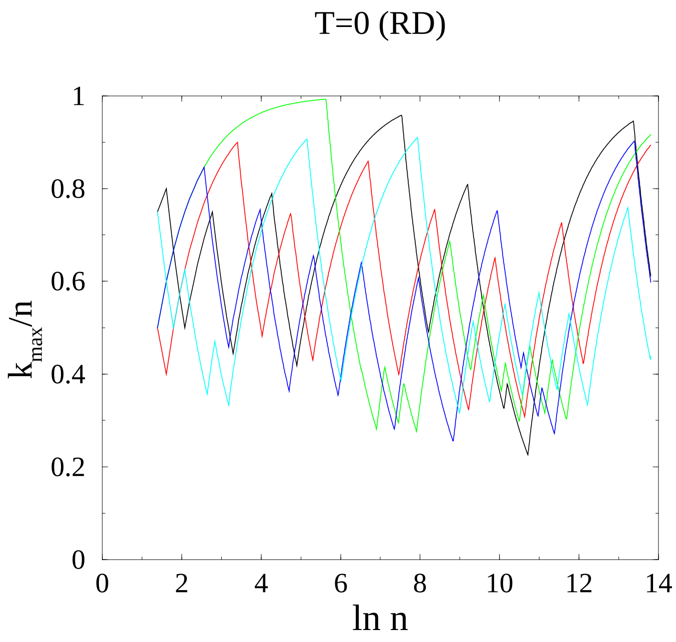
- Multiplicative continuum formalism

  $$N_{m+1} = \frac{N_m}{U_{m+1}} \qquad U_{m+1} \text{ uniform in } [0,1]$$

- *Scale invariance, i.e., stationarity in logarithmic time*

Degree $k_{\max}$ of the leader

Extensive and fluctuating...



Take stroboscopic view of the growth process at the record times N_m

Ratios $R_m = \dfrac{k_{\max}(N_m)}{N_m}$

obey recursion $R_{m+1} = \max(U_{m+1} R_m,\ 1 - U_{m+1})$

This line of thought yields many analytical results

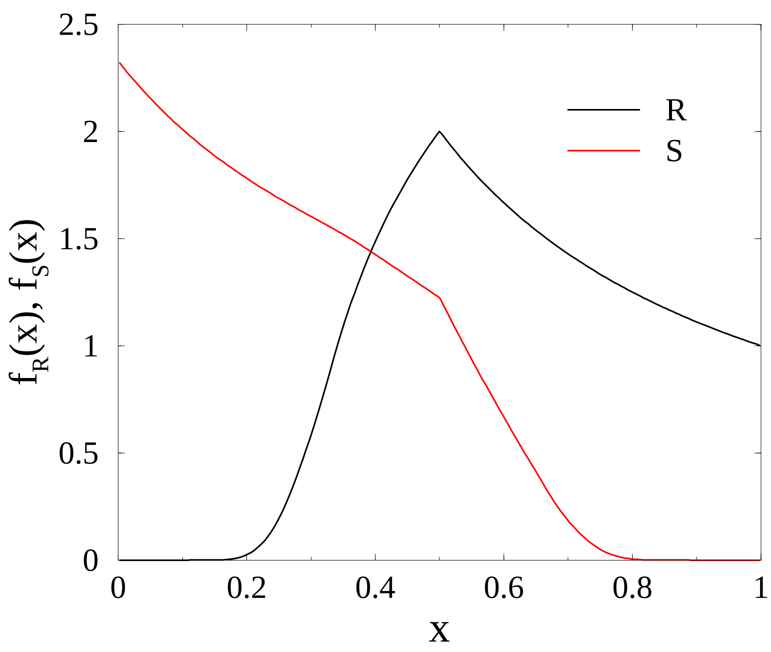
Connection with Kesten variable and 1D disordered systems

A few of our analytical results

- Degree $k_{\max}(n)$ and index $I(n)$ of leader at time n

  Grow linearly in time, as $k_{\max}(n) = Rn$ and $I(n) = Sn$

  Ratios R and S have known limiting distributions



- Probability that a given record node is (will be) a leader

$$\omega = \langle R \rangle = \int_0^\infty \left(1 - \mathrm{e}^{-E(s)}\right) \mathrm{d}s = 0.624329988\ldots$$

$$E(s) = \int_s^\infty \mathrm{e}^{-t}\, \frac{\mathrm{d}t}{t}$$

ω *is the Golomb-Dickman constant... well-known in Combinatorics*

- Number of leads and of distinct leaders up to time n

$$\mathcal{L}(n) = D(n) \approx \omega \ln n$$

Summary of II

Typical history for $T = 0$ (Record-driven growth process)

- *Only record nodes grow*
- *Every record has a chance* ω *to become a leader*
- *All leaders are distinct*
- $\ln n$ *records and* $\omega \ln n$ *leaders up to time* n
- *Degree and index of the leader linear in time and fluctuating*

Results III: T finite (generic BB model)

Reminder: Attachment probability $\Pi_{n\to m} = \dfrac{k_m(n-1)\,q_m}{Z(n-1)}$

Fitnesses are activated: $q_m = \exp(-\varepsilon_m/T)$

Mean-field analysis (Bianconi and Barabási, 2001)

Mean degree of node m at time n knowing its fitness q_m

$$\langle k_m(n)\rangle_{q_m} \approx \left(\frac{n}{m}\right)^{q_m/C}$$

- Growth exponent q_m/C proportional to fitness q_m
- Denominator C obeys *gap equation* $I(T,C) = \displaystyle\int_0^\infty \frac{\rho(\varepsilon)\,d\varepsilon}{C\,e^{\varepsilon/T}-1} = 1$

Condensation transition at finite temperature T_c

$$C = 1 \ \text{for} \ T_c \ \text{s.t.} \ I(T_c, 1) = \int_0^\infty \frac{\rho(\varepsilon)\,d\varepsilon}{e^{\varepsilon/T_c} - 1} = 1$$

- Formal analogy with *Bose-Einstein condensation*
- Condition: $\rho(\varepsilon) \sim \varepsilon^{\theta-1}$ with $\theta > 1$ as $\varepsilon \to 0$ (analogue of $d > 2$ for BEC)
- Condensed phase $(T < T_c)$

  Known condensed fraction $F = 1 - I(T, 1)$

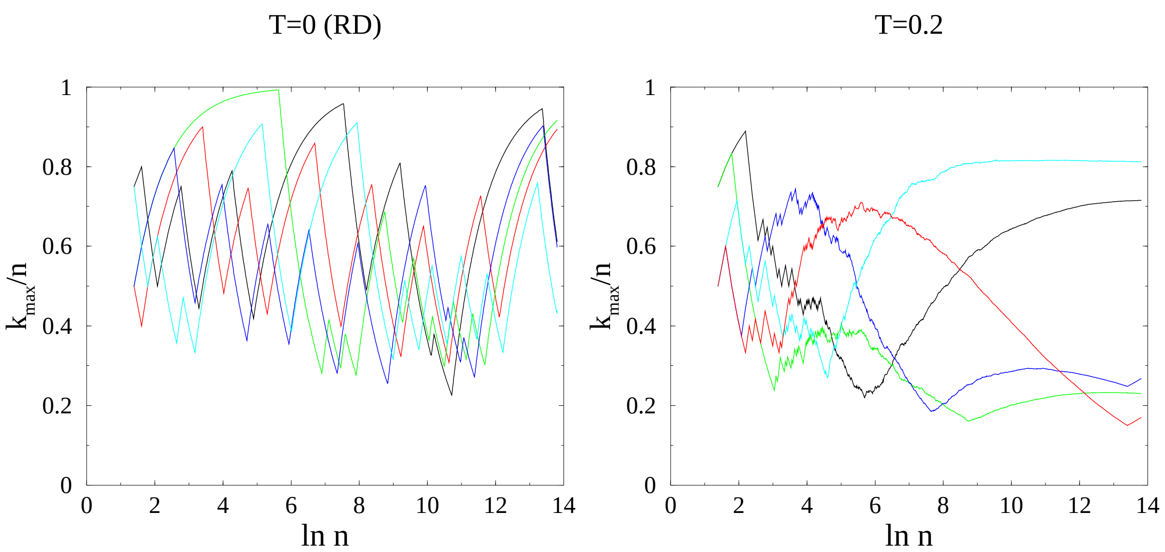
  Degree of condensate (leader) should grow as $k_{\max} \approx F\,n$

What is actually observed in condensed phase?

Numerical study for triangular distribution of energies

$\rho(\varepsilon) = 2\varepsilon$ on $[0,1]$. $\qquad T_c = 0.711716308\ldots$

Degree of leader: $R(n) = k_{\max}(n)/n$



Extensive fluctuations

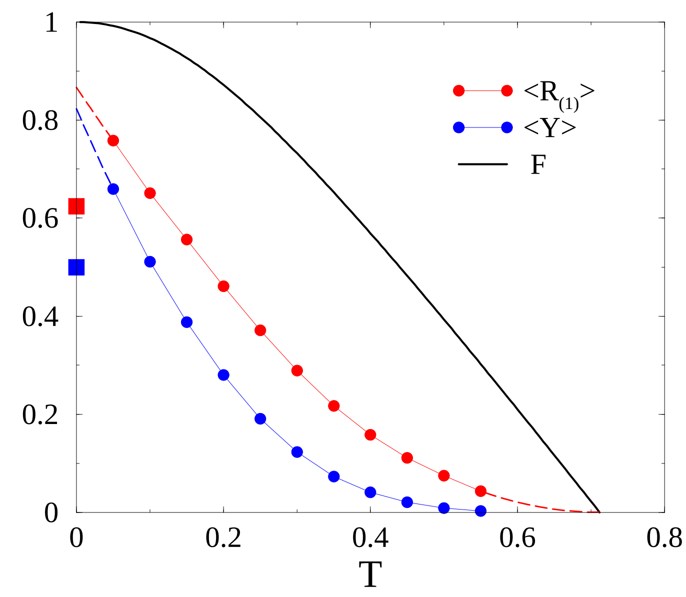
Slow down w.r.t. record-driven case

Far from BEC picture with its constant condensed fraction

Condensation scenario

Infinite hierarchy of fluctuating condensates

- Degrees $k_{(1)} \geq k_{(2)} \geq k_{(3)} \geq \dots$
- Leader with degree $k_{(1)}$ not really singled out
- Ratios $R_{(j)} = k_{(j)}/n$ have non-trivial T -dept joint law
- Statics just provides sum rule $\sum_j R_{(j)} = F$
- Non-trivial higher moments, e.g. $\sum_j R_{(j)}^2 = Y$



History of leaders

The Leader of Today is the Record of Old

Reminder: $\langle k_m(n)\rangle_{q_m} \approx \left(\frac{n}{m}\right)^{q_m/C}$

(1) Original variational extreme-value argument for n.i.i.d. variables

$I \sim ((\ln n)/T)^{\theta}$

(2) Leader is a record with high probability $(\Pi_\infty(T_c) \approx 0.866)$

Hence $D(n) \sim \theta \ln\ln n, \quad S(n) \sim (\ln n)^{-\theta}$

The above holds irrespective of phase: both fluid $(T > T_c)$ and condensed $(T < T_c)$

Crossover time at low temperature: $\tau \sim T^{-\theta}$

Take home messages

1. The fate of leaders

A more efficient search engine is launched today. Will it outdo Google? If so, when?

The Leader of Today is the Record of Old

- Yes with high probability. But in a very distant *(exponentially large)* future
- Beware of broad crossovers to other regimes *(BA, RD)*

2. Nonequilibrium condensation dynamics

Condensed phase of BB model provides exotic scenario (w.r.t. BEC & ZRP)

- Infinite hierarchy of fluctuating condensates
- Very slow turnover