Introduction to network dynamics

Marta Arias, Ramon Ferrer-i-Cancho, Argimiro Arratia

Universitat Politècnica de Catalunya

Complex and Social Networks (2022-2023)
Master in Innovation and Research in Informatics (MIRI)

Instructors

- Ramon Ferrer-i-Cancho, rferrericancho@cs.upc.edu, `http://www.cs.upc.edu/~rferrericancho/`
- Argimiro Arratia, argimiro@cs.upc.edu, `http://www.cs.upc.edu/~argimiro/`

Please go to http://www.cs.upc.edu/~csn for all course's material, schedule, lab work, etc.

Models that generate networks [Caldarelli, 2007]

- The Barabási-Albert model (growth and preferential attachment).
- Copying models
- Fitness based model
- Optimization models

Each model produces a network through different dynamical principles/rules.

The Barabási-Albert model

Example from citation networks, where $p(k) \sim k^{-3}$ [Redner, 1998].

The evolution of an undirected network over time t.

1. $t = 0$, a disconnected set of n_0 vertices (no edges).
2. At time $t > 0$, add a new vertex with m_0 edges:
   - The new vertex connects to the i-th vertex with probability

$$\pi(k_i) = \frac{k_i}{\sum_j k_j}$$

Thus

$$n = n_0 + t$$

$$m = \frac{1}{2}\sum_{j=1}^{n} k_i = m_0 t$$

The growth of a vertex degree over time I

The dependence of k_i on time

- Treat k_i as a continuous variable (although it is not).
- The variation of degree over time (on average) is

$$\frac{\partial k_i}{\partial t} = m_0 \pi(k_i) = m_0 \frac{k_i}{2 m_0 t} = \frac{k_i}{2t}$$

- t_i is the time at which the i-th vertex was introduced.
- m_0 is the degree of the i-th vertex at time t_i.
- Integrate on both sides of

$$\frac{\partial k_i}{k_i} = \frac{\partial t}{2t} \rightarrow \int_{m_0}^{k_i} \frac{\partial k_i}{k_i} = \frac{1}{2} \int_{t_i}^{t} \frac{\partial t}{t}$$

The growth of a vertex degree over time II

Finally,

$$k_i(t) \approx m_0 \left(\frac{t}{t_i}\right)^{1/2}$$

A non-rigorous proof that $p(k) \approx k^{-3}$ I

Sketch of the proof [Barabási et al., 1999]

- Starting point: $k_i(t) = m_0 \left(\frac{t}{t_i}\right)^{1/2}$
- Final goal: obtain $p(k)$ through

$$p(k) \approx \frac{\partial p(k_i < k)}{\partial k}$$

- Intermediate goal: calculate $p(k_i < k)$

A rigorous proof is available [Bollobás et al., 2001]

A non-rigorous proof that $p(k) \approx k^{-3}$ II

- $p(k_i < k)$: the probability that the i-th vertex has degree lower than k.
-
$$p(k_i < k) = p\left(m_0 \left(\frac{t}{t_i}\right)^{1/2} < k\right) = p\left(t_i > \frac{m_0^2 t}{k^2}\right)$$
- $p(t_i = \tau) = 1/(n_0 + t)$ for $n_0 = 1$ (for $t_i \leq \tau$).
- $p(t_i = \tau) \approx 1/(n_0 + t)$ for $n_0 > 1$ but small.

$$p\left(t_i > \frac{m_0^2 t}{k^2}\right) = 1 - p\left(t_i \leq \frac{m_0^2 t}{k^2}\right) = 1 - \sum_{\tau=0}^{\frac{m_0^2 t}{k^2}} p(t_i = \tau)$$

A non-rigorous proof that $p(k) \approx k^{-3}$ III

- $$p\left(t_i > \frac{m_0^2 t}{k^2}\right) \approx 1 - \frac{m_0^2 t}{n_0 + t} k^{-2}$$
- $$p(k) \approx \frac{\partial p(k_i < k)}{\partial k} \approx \frac{2 m_0^2 t}{n_0 + t} k^{-3}$$
- $p(k) \approx c k^{-\gamma}$ with $\gamma = 3$ and $c = \frac{2m_0^2 t}{n_0 + t}$.

More rigorous proofs are available [Newman, 2018].
Exercise: a more precise calculation for $p(t_i = \tau)$.

Deeper thinking

- $m_0 \leq n_0$ is needed.
- Initial conditions: if there are n_0 disconnected vertices, then $\pi(k_i)$ is undefined initially. Solutions:
  - Another initial condition, e.g., a complete graph of n_0 nodes.
  - Same initial condition but different preferential attachment rule, e.g.,

$$\pi(k_i) = \frac{k_i + 1}{\sum_j (k_j + 1)}$$

- Some limitations:
  - Global knowledge is required by π.
  - $p(k) \sim k^{-\gamma}$ with $\gamma = 3$ is suitable for article citation networks [Redner, 1998] but $\gamma < 3$ in many real networks, e.g., global syntactic dependency networks (lab session and [Ferrer-i Cancho et al., 2004]).

The origins of the power-law in Barabási-Albert model I

Controlling for the role of growth and preferential attachment [Barabási et al., 1999]

- Hypothesis: preferential attachment is vital for obtaining a power-law (in that model)
- Test: Replacing the preferential attachment by uniform attachment (all vertices are equally likely) $\rightarrow p(k) = ae^{-ck}$.
- Hypothesis: growth is vital for obtaining a power-law (in that model)
- Test: suppressing growth: fixed number vertices $\rightarrow k$ follows a "Gausian" distribution.

The origins of the power-law in Barabási-Albert model II

Controlling for the hidden assumptions of the preferential attachment rule

- Generalizing the preferential attachment [Krapivsky et al., 2000]

$$\pi(k_i) = \frac{k_i^\delta}{\sum_j k_j^\delta}$$

  - $\delta = 1 \rightarrow$ original B.A. model.
  - $\delta > 1 \rightarrow$ one node dominates (very pronounced effect for $\delta > 2$).
  - $\delta < 1 \rightarrow$ combination of power-law with stretched exponential.

The effect of replacing preferential attachment by random attachment

The growth of a vertex degree over time

- Recall $n(t) = n_0 + t$.
- The variation of degree over time (on average) is

$$\frac{\partial k_i}{\partial t} = \frac{m_0}{n(t-1)}$$

- Integrate on both sides of

$$\partial k_i = m_0 \frac{\partial t}{n(t-1)} \rightarrow \int_{m_0}^{k_i} \partial k_i = m_0 \int_{t_i}^{t} \frac{\partial t}{n(t-1)}$$

The effect of replacing preferential by random attachment

Finally,

$$\begin{aligned} k_i(t) &\approx m_0 \left(\log \frac{n(t-1)}{n(t_i-1)} + 1\right) \\ &= m_0 \left(\log \frac{n_0+t-1}{n_0+t_i-1} + 1\right) \end{aligned}$$

A non-rigorous proof that $p(k) \sim e^{k/m_0}$ I

Sketch of the proof [Barabási et al., 1999]

- Starting point: $k_i(t) = m_0 \left(\log \frac{n_0+t-1}{n_0+t_i-1} + 1 \right)$
- Final goal: obtain $p(k)$ through

$$p(k) \approx \frac{\partial p(k_i < k)}{\partial k}$$

- Intermediate goal: calculate $p(k_i < k)$

A non-rigorous proof that $p(k) \sim e^{k/m_0}$ II

- $p(k_i < k)$: the probability that the i-th vertex has degree lower than k.
- $$\begin{aligned} p(k_i < k) &= p\left(m_0\left(\log \frac{n_0+t-1}{n_0+t_i-1}+1\right) < k\right) \\ &= p\left(t_i > (n_0+t-1)e^{1-\frac{k}{m_0}} - n_0 + 1\right) \end{aligned}$$
- Recall $p(t_i = \tau) \approx 1/(n_0 + t)$ for $n_0 > 1$ but small.

$$p\left(t_i > ...\right) = 1 - p\left(t_i \leq ...\right) = 1 - \sum_{\tau=0}^{...} p(t_i = \tau)$$

A non-rigorous proof that $p(k) \sim e^{k/m_0}$ III

- $$p\left(t_i > \ldots\right) \approx 1 - \frac{1}{n_0 + t} \ldots$$
- Then,
$$p(k_i < k) = p(t_i > \ldots) = 1 - \frac{1}{n_0 + t}\left((n_0 + t - 1)e^{1 - \frac{k}{m_0}} - n_0 + 1\right)$$
- Finally (for long times)
$$p(k_i < k) = 1 - e^{1 - \frac{k}{m_0}}$$
- $$p(k) \approx \frac{\partial p(k_i < k)}{\partial k} \approx \frac{e}{m_0} e^{-\frac{k}{m_0}}$$
- $p(k) \approx Be^{-\beta k}$ with $B = e/m_0$ and $\beta = 1/m_0$.

The effect of suppressing vertex growth

The new vertex is replaced by a vertex chosen uniformly at random.

Evolution of the degree distribution as t increases [Barabási et al., 1999]

- Initial phase: power-law.
- Intermediate phase: Gausian-like.
- Final state (complete graph): $\delta_{n_0-1,k}$ (Kronecker's delta function).

Copying vertices

Motivation:

- Producing a new web page by copying another web page and making some modifications (some of the hyperlinks may remain while new hyperlinks may be added).
- Protein interaction networks [Vázquez et al., 2003]. Genetic evolution: duplication of DNA + mutations may produce new proteins that inherit some interaction properties from the original protein.

Features:

- Network growth (new vertices) + copying + rewiring.
- Local rule (no global knowledge, the degree of all vertices).

The copying model I

- Start with some initial configuration.
- At every time-step: a the vertex is chosen uniformly at random).
  - **Duplication**: the vertex is duplicated to produce a new vertex (the new vertex has out-degree m_0).
  - **Divergence**: each out-going connection is rewired with probability α or kept with probability $1 - \alpha$.
    - Rewiring means changing the end-point by a vertex chosen uniformly at random.

The copying model II

- Here: simple copying model [Caldarelli, 2007].
  - Directed network. Every new vertex sends m_0 edges to old vertices.
  - For vertices added at time $t > 0$, out-degree is constant (m_0) while in-degree varies.
- Other versions of the copying model with more or different parameters
  [Vazquez et al., 2003, Pastor-Satorras et al., 2003].

The mathematical properties of a copying model I

$$\frac{\partial k_i^{in}(t)}{\partial t} = \frac{1-\alpha}{N} k_i^{in}(t) + m_0 \frac{\alpha}{N},$$

where

- $\frac{1-\alpha}{N} k_i^{in}(t)$ is the contribution from retained edges of a vertex pointing to vertex i that is duplicated.
- $m_0 \frac{\alpha}{N}$ is the contribution from rewired edges of the duplicated vertex (the expected number of times that the i-th is hit in those rewirings).
- $N \approx t$ (linearly growing network)
- Warning: wild assumptions about $\frac{\partial k_i^{in}(t)}{\partial t}$ are being made and thus numerical calculations to check the analytical results are needed.

The mathematical properties of a copying model II

- $$k_i^{in}(t) = \frac{m_0\alpha}{1-\alpha}\left[\left(\frac{t}{t_i}\right)^{1/2} - 1\right]$$
- t_i: arrival time of the i-th vertex.
- $$p(k^{in}) \sim \left[k^{in} + \frac{m_0\alpha}{1-\alpha}\right]^{-\frac{2-\alpha}{1-\alpha}}$$
- $p(k^{in}) \sim k^{-2}$ for $\alpha = 0$

The copying model versus the Barabási-Albert model

Nice properties:

- Emergence of the preferential attachment rule from local principles! (the original preferential attachment is a global principle)
- A wider and more realistic range of exponents is captured!

Connecting according to vertex fitness (not vertex degree)

- An alternative to preferential attachment, e.g., when the degree of other vertices is not available to newcomers.
- Linking according to intrinsic properties (that determine the *fitness* of a vertex)
  - Authoritativeness, social success or status, scientific relevance, interaction strength (of the vertex).

A general fitness model [Caldarelli et al., 2002]

- Setup: start with N vertices.
- Fitness: assign to every vertex a fitness.
  - x_i is the fitness of the i-th vertex.
  - The fitness of a vertex is obtained producing a random number following the probability density function $\rho(x)$ (harder calculations with a probability mass function)
- Linkage: for every couple of vertices i and j, draw an edge with a probability given by a linking function $f(x_i, x_j)$ (in undirected networks, f is symmetric, $f(x_i, x_j) = f(x_j, x_i)$).

Comments:

- A generalization of the Ërdos-Rényi model, where $f(x_i, x_j) = p$.
- Reminiscent of the network *configuration model*.

Degree distribution in a fitness model I

- The degree distribution is not necessarily a power law (e.g., $f(x_i, x_j) = p$).
- Consider $f(x_i, x_j) = (x_i x_j)/x_M^2$ where x_M is the largest value of x in the network. Then the mean degree of a node of fitness x is

$$\begin{aligned} k(x) &= N \int_0^\infty f(x,y)\rho(y)dy \\ &= \frac{N\ x}{x_M^2}\int_0^\infty y\rho(y)dy = \frac{N\langle x\rangle}{x_M^2}x. \end{aligned} \tag{1}$$

and

$$p(k) = \frac{x_M^2}{N\langle x\rangle}\rho\left(\frac{x_M^2}{N\langle x\rangle}k\right) \tag{2}$$

Degree distribution in a fitness model II

- If fitness follows a power law, i.e.

$$\rho(x) \sim x^{-\beta} \tag{3}$$

  then $p(k) \sim k^{-\beta}$ [Caldarelli et al., 2002]
- Motivation: Zipf's law: $p(x) \sim x^{-\beta}$ in many contexts (word frequencies, population size of cities...).

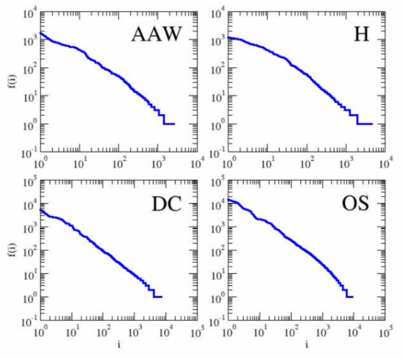
George Kingsley Zipf

- The founder of modern quantitative linguistics.
- Interested in unifying principles of nature (principle of least effort).

- Zipf, G. K. (1949) Human Behavior and the Principle of Least Effort. Addison-Wesley.
- Zipf. G. K. (1935) The Psychobiology of Language. Houghton-Mifflin.

Zipf's law

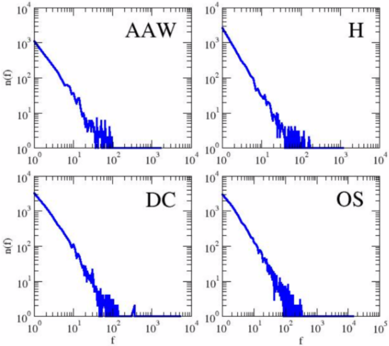
The rank histogram (number-rank)



- Empirical law [Zipf, 1949].
- Apparently universal.
- Popularized but not discovered by G. K. Zipf
- $n(i) \sim i^{-\alpha}$
- $\alpha \approx 1$

Zipf's law: a less popular version

The frequency histogram (number-frequency)



- Less popular than the rank histogram.
- $n(f) \sim f^{-\beta}$
- $\beta \approx 2$
- $\beta = 1/\alpha + 1$

Degree distribution in a fitness model III

- If fitness is not power-law distributed, it is still possible to obtain a power-law distributed degrees [Caldarelli et al., 2002].
- Example:
  - $\rho(x) = e^{-x}$ (probability density function $\rho(x) = \lambda e^{-\lambda x}$ with $\lambda = 1$, $x \geq 0$)
  - $f(x_i, x_j) = \theta(x_i + x_j - z)$ where
    - z is a threshold parameter
    - $\theta(x)$ is the Heaviside function, i.e.

$$\theta(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$$

  - $p(k) \sim k^{-2}$
  - Generalization $f(x_i, x_j) = \theta(x_i^a + x_j^a - z^a)$ being a an integer, still $p(k) \approx k^{-2}$ (logarithmic corrections might be necessary).

Optimization in a network

Desired properties of a network:

- Small geodesic distance.
- Small number of edges (edge = cost).

Trade-off between both:

- Smallest geodesic distance: complete graph.
- Smallest number of links: tree (but a linear tree has the largest distance possible).

The energy function to minimize II

Two normalized metrics

- $\rho = \langle k \rangle / (N-1)$ (density of an undirected network without loops)
- $\Delta = d/d_{linear}$ with $d_{linear} = (N+1)/3$ (do you remember $(N+1)/3$ somewhere else?)

Networks that minimize

$$E(\lambda) = \lambda \Delta + (1-\lambda)\rho$$

with the the following constraints:

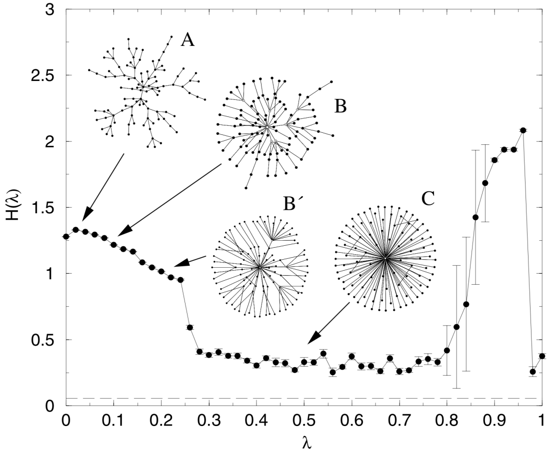
- The network size (in vertices) is constant.
- The network has to remain connected.

The energy function to minimize II

$$E(\lambda) = \lambda\Delta + (1-\lambda)\rho$$

- $\lambda = 0$: only the number of links is minimized.
- $\lambda = 1$: only the geodesic distances are minimized.
- Networks with exponential and power-law degree distribution appear in between.
- See Fig. 7.4 [Ferrer-i Cancho and Solé, 2003].

Figure 7.4 [Ferrer-i Cancho and Solé, 2003]



Further comments

- $E(\lambda)$ is reminiscent of $AIC = -\log L + 2K$.
- The regimes in Fig. 7.4 [Ferrer-i Cancho and Solé, 2003] are reminiscent of those of a generalized BA model [Krapivsky et al., 2000]. Is there some equivalence between both (λ vs δ)?
- Future work: remove the connectedness constraint. How?

References I

Barabási, A.-L., Albert, R., and Jeong, H. (1999).
Mean-field theory for scale-free random networks.
Physica A: Statistical Mechanics and its Applications, 272(1-2):173–187.

Bollobás, B. e., Riordan, O., Spencer, J., and Tusnády, G. (2001).
The degree sequence of a scale-free random graph process.
Random Structures & Algorithms, 18(3):279–290.

Caldarelli, G. (2007).
Scale-free networks: complex webs in nature and technology.
Oxford University Press.

References II

Caldarelli, G., Capocci, A., De Los Rios, P., and Munoz, M. A. (2002).
Scale-free networks from varying vertex intrinsic fitness.
Physical review letters, 89(25):258702.

Ferrer-i Cancho, R. and Solé, R. V. (2003).
Optimization in complex networks.
In *Statistical mechanics of complex networks*, pages 114–126. Springer.

Ferrer-i Cancho, R., Solé, R. V., and Köhler, R. (2004).
Patterns in syntactic dependency networks.
Physical Review E, 69(5):051915.

References III

Krapivsky, P. L., Redner, S., and Leyvraz, F. (2000).
Connectivity of growing random networks.
Physical review letters, 85(21):4629.

Newman, M. (2018).
Networks.
Oxford university press.

Pastor-Satorras, R., Smith, E., and Solé, R. V. (2003).
Evolving protein interaction networks through gene duplication.
Journal of Theoretical biology, 222(2):199–210.

References IV

Redner, S. (1998).
How popular is your paper? an empirical study of the citation distribution.
The European Physical Journal B-Condensed Matter and Complex Systems, 4(2):131–134.

Vazquez, A., Flammini, A., Maritan, A., and Vespignani, A. (2003).
Global protein function prediction from protein-protein interaction networks.
Nature biotechnology, 21(6):697.

References V

Vázquez, A., Flammini, A., Maritan, A., and Vespignani, A. (2003).
Modeling of protein interaction networks.
Complexus, 1(1):38–44.

Zipf, G. K. (1949).
Human behavior and the principle of least effort: An introduction to human ecology.
Addison Wesley.