Network Models I.: Random Networks & Small Worlds

IV124

Josef Spurný & Eva Výtvarová

Faculty of Informatics, Masaryk University

March 17, 2023

Random Graph Model

Why model a random graph:

- Properties can be mathematically derived
- Useful for comparison with a real network:
  - What are the differences?
  - What does it tell us about the network?

Erdös-Rényi random graph model:

- $G(N, L)$ model, where L is a number of links randomly placed among N nodes (proposed by Erdös and Rényi)
- $G(N, p)$, where N is the number of nodes and p is the probability of connection between two nodes (more commonly used, but actually proposed by Edgar Gilbert in 1959)

Erdös-Rényi Model: Properties

A probability, that a random network has exactly $|E|$ edges, is defined by binomial distribution:

- $P(|E|) = \binom{E_{max}}{|E|} p^{|E|} (1-p)^{E_{max}-|E|}$
- where $E_{max} = N(N-1)/2$ is the maximum number of edges

A probability that a randomly selected node has a degree k:

- Binomial distribution: $P(k) = \binom{N-1}{k} p^k (1-p)^{N-1-k}$
  - $\binom{N-1}{k}$ selection of k nodes
  - p^k: probability of k edges forming
  - $(1-p)^{N-1-k}$: absence of remaining edges
  - $\overline{k} = p(N-1)$

Erdös-Rényi Model: CC

Clustering coefficient

- $C_i = \frac{L_i}{k_i(k_i-1)}$
- substituting L_i with $p\frac{k_i(k_i-1)}{2}$ – probability of a link between neighbors
- hence $C_i = \frac{pk_i(k_i-1)}{k_i(k_i-1)} = p$

For real sparse networks, $\overline{C}$ is indeed very small.

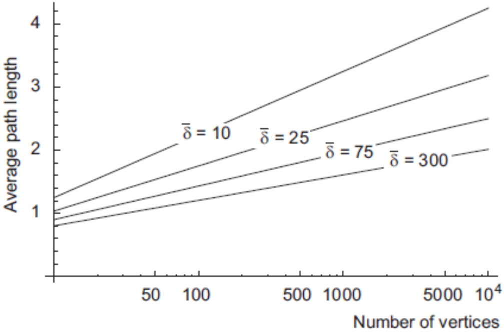
Erdös-Rényi Model: Average Path Length

Derivation

- consider a network with a given $\overline{k}$
- on average, a node has $\overline{k}^d$ neighbors at a distance of d
- thus, the number of nodes at a distance of d is $N(d) = \frac{\overline{k}^{d+1}-1}{\overline{k}-1}$
- but $N(d) \leq N$, so $\overline{k}^{d_{max}} \approx N$ and $d_{max} = \frac{log(N)}{log(\overline{k})}$
- for most networks, a good approximation for the average path length is $\overline{d} \approx \frac{ln(N)}{ln(\overline{k})}$

Erdös-Rényi Model: Average Path Length

$\overline{\delta} = \overline{k} =$ average degree



Gephi demo: Random Graphs

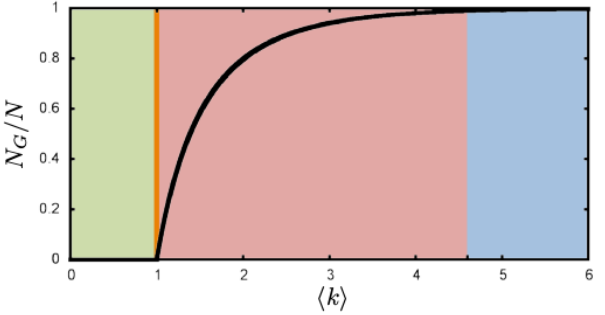
...

Netlogo demo: Random Graphs

...

Giant Component

N_G = size of the largest component



Giant Component

Subcritical regime $\overline{k} < 1$

- no giant component
- largest clusters $N_G \approx \ln N$
- clusters are trees of comparable size, there is no "winner"
- clusters grow much slower than network, hence $N_G/N \to 0$ as $N \to \infty$

Giant Component

Critical point $\overline{k} = 1$

- no giant component, numerous small components
- largest clusters are typically much larger than in subcritical regime, $N_G \approx N^{2/3}$
- still, the largest cluster connects only an insignificant fraction of all nodes
- clusters may contain loops

Giant Component

Supercritical regime $\overline{k} > 1$

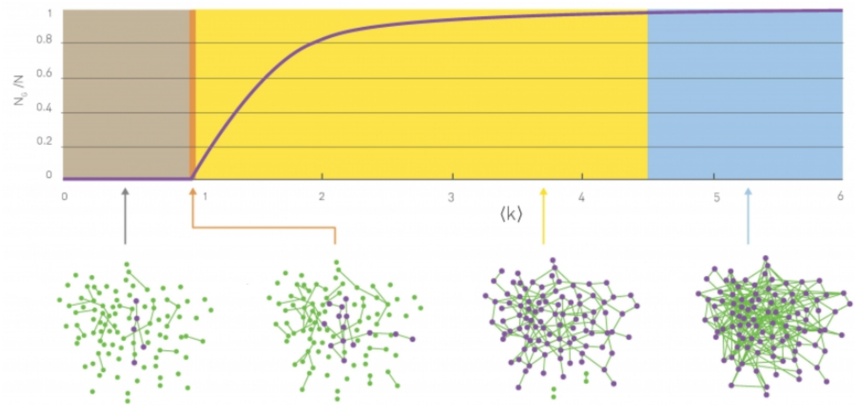
- one giant component
- relevant to most real-world systems
- largest clusters $N_G \approx (p - p_c)N$, where p_c is $\frac{1}{N}$
- small clusters are trees (isolated vertices)

Giant Component

Fully connected regime $\bar{k} \geq \ln N$

- one giant component
- giant component absorbs all nodes and clusters, hence $N_G = N$ (no isolated nodes)
- yet the network is still relatively sparse
- we receive complete graph only when $\bar{k} = N - 1$

Giant Component Evolution

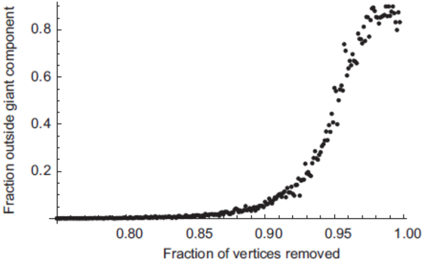


Netlogo demo: Component

...

Giant Component: Resistance to Node Failure

ER(2000, 0.015)



We may need to remove up to 70% nodes before network partitions. Removing 95% nodes, half of the remaining are still connected through a path.

Strength of Weak Links[1]

Research question: How do people seek a new job?

- Hypothesis: Your family would help you
- Study results: Most commonly, a friend of a friend will give you a good tip

[1] Granovetter, M. S. (1973). The strength of weak ties.

Small World Problem[2]

What is the probability that two randomly selected people will know each other?

- 300 individuals from different places in the USA
- the goal was to deliver a letter to a target person in Boston through personal contacts

Results:

- 64 successful chains
- on average 6.2 steps: 6 degrees of separation

[2] Milgram, S. (1967). The small world problem. *Psychology today*, 2(1), 60-67.

Milgram's Experiment: Why as low as 6?

Random social networks:

- assumes 500-1500 contacts per person[3]
- for a random network, three steps involve $\sim 500^3 = 125 \cdot 10^6$ individuals

Small World Property: $\overline{d} \approx \frac{ln(N)}{ln(\overline{k})}$

- for US pop $\approx 330M$ and 500 contacts, $\overline{d} \approx 3.16$ (EU pop $\approx 450M$, then $\overline{d} \approx 3.20$)
- in general, $ln(N) \ll N$, therefore path length is of orders of magnitude smaller than network size
- *small world phenomenon depends logarithmically on network size*

[3] Pool & Kochen (1978)

Which Model to Choose?

Desired properties:

- small diameter (average path length): $l \approx \ln(N)$
- high clustering coefficient: $C \gg C_{rand}$

model	clusters	small diameter
Erdös Rényi	no	yes
Barabási-Albert[4]	no	yes
grid	yes	no
?	yes	yes

[4] this model did not yet exist at the time

Watts-Strogatz Model[5]

$WS(N, k, p)$

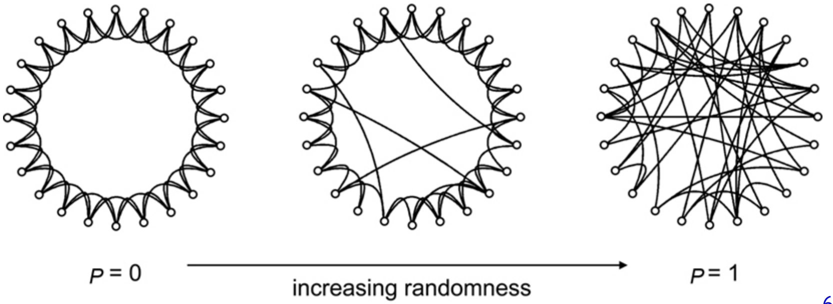
Procedure:

- start with N nodes connected to their k nearest neighbors
- for each edge, with probability p, randomly rewire the target node

For certain values of p, we obtain both high C and low l

[5] Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world'networks. Nature, 393(6684), 440-442.

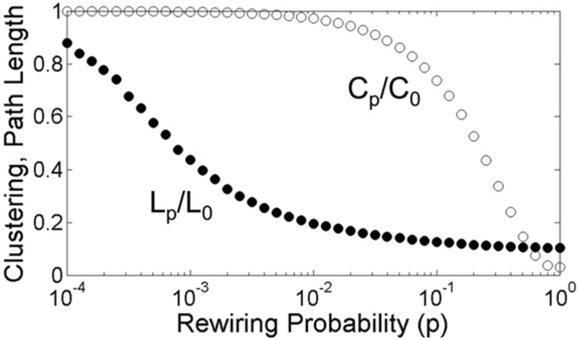
Watts-Strogatz Model



[6]

[6] Sporns O. (2011)

Watts-Strogatz Model



[7]

http://bit.ly/1O2WBIL

[7] Sporns O. (2011)

Watts-Strogatz: ukázka

...

Forest Fire Model[8]

Motivation:

- Watts-Strogatz model does not create a scale-free network
- Random rewiring is difficult to interpret

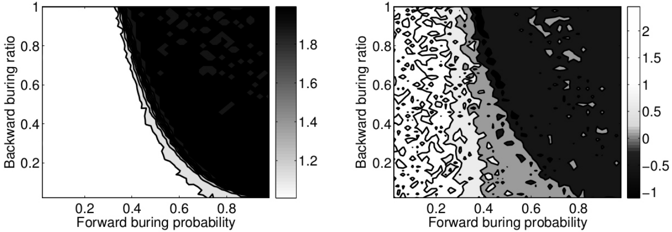
Procedure:

- At each step, we add a node u
- We randomly select and *ignite* a connection point
- The fire iteratively spreads with probability p, r times less likely through incoming edges
- We attach the burned edges to u

[8] Leskovec et. al (2005)

Forest Fire Model – Properties

- Preferential attachment: power-law degree distribution
- Community-driven attachment: clustering
- Only two parameters



Densification exponent   Diameter [9]

[9] Leskovec et. al (2005)

Small Worlds vs Efficiency[10]

Intuition:

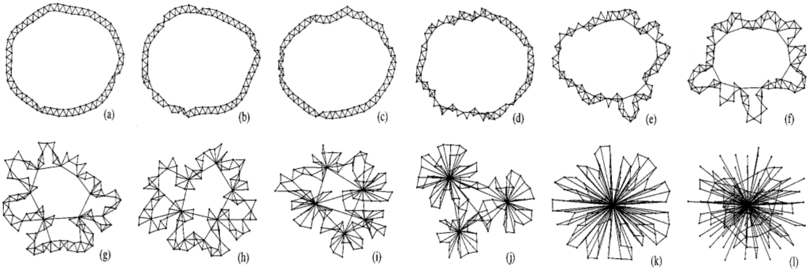
- In physical networks (transportation, neural, etc.), there is a trade-off between maximum connectivity and the cost of building connections
- The evaluation function $E = \lambda L + (1 - \lambda) W$
- L is the characteristic path length, W is the total cost of connections (for a single edge, it depends on the distance between nodes), λ indicates a preference for L vs W

[10] Mathias & Gopal (2000)

Small worlds vs efficiency

Result:

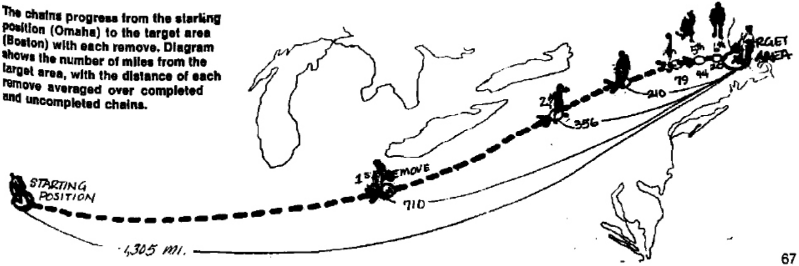
- When network minimizes wiring ($\lambda \to 0$), regular graphs are obtained
- When network maximizes wiring ($\lambda \to 1$), random networks are obtained
- For intermediate values, we get small worlds with **hubs** – note that hubs do not emerge in classic WS model



[11]

[11] Mathias & Gopal (2000). Small Worlds: How and Why

Milgram’s Experiment – Navigation



Observation:

- with each step, the letters get closer to their addressee

Milgram's Experiment Nowadays[12]

Navigation over geo-tagged Twitter Network:

- knowing only location, it is quite easy to reach the right city
- however, on a smaller scale (inside the city), a letter gets 'lost' and spends much more time looking for the right addressee
- In Milgram's experiment, participants used other than just geographic info (e.g., occupation, social status...)

[12] Szüle et al. (2014). Lost in the City: Revisiting Milgram's Experiment in the Age of Social Networks.

Kleinberg's Model[13]

Motivation:

- Utilize local knowledge of the geographic location of the target and other nodes

Procedure:

- Nodes are placed on a grid
- Random edges are added:
  $p(u, v) = d(u, v)^{-\alpha}$
  where $\alpha > 0$

Findings:

- there is a specific value of α which allows optimal (fast) navigation ($\alpha = 2$)
- any other value requires asymptotically larger delivery time

[13] Kleinberg, J. (2000). Navigation in a small world.

MUNI
FACULTY
OF INFORMATICS