Network Science Notes

Wei-Hsiang Sun

July, 2020

NETWORK SCIENCE INTRODUCTION

1. Vulnerability Due to Interconnectivity
2. Two Forces that Helped Network Science
   - The Emergence of Network Maps
   - The Universality of Network Characteristics
3. The Characteristics of Network Science
   - Interdisciplinary Nature
   - Quantitative and Mathematical Nature
   - Computational Nature
4. Societal Impact
5. Scientific Impact

NETWORK SCIENCE GRAPH THEORY

1. The Bridges of Konisgbergs
2. Networks and Graphs
   - Number of nodes/vertices: N
   - Number of links/edges: L
   - Links can be directed or undirected,
     If directed, it is also called a Digraph
3. Degree, Average Degree, and Degree Distributions
   - Degree of a node: number of links it has to other nodes
   - for undirected graph:
     - $L = \frac{1}{2}\sum_{i=1}^{N} k_i$
     - $\langle k \rangle = \frac{1}{N}\sum_{i=1}^{N} k_i = \frac{2L}{N}$

     for directed graph:
     - $k_i = k_i^{in} + k_i^{out}$
     - $L = \sum_{i=1}^{N} k_i^{in} = \sum_{i=1}^{N} k_i^{out}$
     - $\langle k^{in} \rangle = \frac{1}{N}\sum_{i=1}^{N} k_i^{in} = \langle k^{out} \rangle = \frac{1}{N}\sum_{i=1}^{N} k_i^{out} = \frac{L}{N}$
   - Degree Distribution p_k
     $p_k = \frac{N_k}{N}$
     $\langle k \rangle = \sum_{k=0}^{\infty} k p_k$

4. Adjacency Matrix
   - $A_{ij} = 1$ if there is a link pointing from node j to i
     $A_{ij} = 0$ if nodes i and j are not connected to each other
   - for undirected network
     $k_i = \sum_{j=1}^{N} A_{ji} = \sum_{j=1}^{N} A_{ij}$
     for digraphs $k_i^{in} = \sum_{j=1}^{N} A_{ij}, \quad k_i^{out} = \sum_{j=1}^{N} A_{ji}$
5. Real Networks are Sparse
   for undirected network:
   $L_{max} = \genfrac{}{}{0pt}{}{N}{2} = \frac{N(N-1)}{2}$
6. Weighted Networks
   $A_{ij} = w_{ij}$, elements carries weight
7. Bipartite Networks
   definition: a network whose node can be divided into two disjoint sets U and V such that each link connects a U node to a V node
   • Projections
   Projection U: connects two U nodes by a link if they are linked to the same V-nodes in the bigraph
   Projection V: connects two V nodes by a link if they are linked to the same U-nodes in the bigraph

8. Paths and Distances

- Path Length: the route that runs along the links of the network
- Shortest Path: the path with the fewest number of links, it is also known as distance between i and j, denoted by d_{ij}
- Network Diameter: denoted by d_{max}, is the maximum shortest path in the network, can be found through BFS algorithm
- Average Path Length: (for directed)

$$d = \frac{1}{N(N-1)} \sum_{\substack{i,j=1,N \\ i \neq j}} d_{i,j}$$

- Cycle (path with the same start and end node)
  Eulerian path (path that goes through link once),
  Hamiltonian Path (path that goes through node once)
- Number of Shortest Path Between Nodes
  Let number of distance d paths between i and j be denoted by $N_{ij}^{(d)}$, then $N_{ij}^{(d)} = A_{ij}^{d}$

9. Connectedness
   - situation where $d_{ij} = \infty$
   - existence of subnetworks components (clusters)
   - bridge: link that connects disconnected components
   - ways of finding diconnectedness: 1. matrix blocks
     2. BFS algorithm
10. Clustering Coeffficient
   - Clustering Coefficient: for a node i with degree k_i,it is

   defined as:

$$C_i = \frac{2\langle L_i \rangle}{k_i(k_i - 1)}$$

   where L_i represents the number of links between the k_i neighbors of node i, $L_i \sim b(\frac{k_i(k_i-1)}{2}, p)$.
   - Average Clustering coefficient

   captures the average of C_i over all nodes, is defined by:

$$\langle C \rangle = \frac{1}{N} \sum_{i=1}^{N} C_i$$

11. Common Network Characteristics
    - Undirected
    - Self-loops: i's where $A_{ii} = 1$
    - Multigraph: multiple links to nodes
    - Directed
    - Weigted
    - Complete Graph(Clique): all nodes are connected to each other

RANDOM NETWORKS

1. Introduction
2. The Random Network Model
   Core idea: a random network consists of N nodes where each node pair is connected with probability p
   - $G(N, L)$ Model: L randomly placed links
   - $G(N, p)$ Model
3. Number of Links

$$
\begin{aligned}
L &\sim b\left(\frac{N(N-1)}{2}, p\right) \\
p_L &= \binom{\frac{N(N-1)}{2}}{L} p^L (1-p)^{\frac{N(N-1)}{2}-L} \\
\langle k \rangle &= \frac{2\langle L \rangle}{N} = p(N-1)
\end{aligned}
$$

4. Degree Distribution
   - Binomial Distribution

$$k \sim b(N-1, p)$$
$$p_k = \binom{N-1}{k} p^k (1-p)^{N-1-k}$$

   - Poisson Distribution

$$k \sim Poisson(\langle k \rangle)$$
$$p_k = \frac{\langle k \rangle^k e^{-\langle k \rangle}}{k!}$$

5. Real Networks are Not Poisson
   - Poisson underestimate the number of high degree nodes
   - Spread of real world network is much higher
   - the reason for lacking of large hubs

     since $k! \sim [\sqrt{2\pi k}] \left(\frac{k}{e}\right)^k$

     thus $p_k = \frac{e^{-\langle k \rangle}}{\sqrt{2\pi k}} \left(\frac{e\langle k \rangle}{k}\right)^k$,where p_k decreases rapidly as $k \uparrow$

6. The Evolution of A Random Network
   - p gradually increases with time
   - let N_G be the size of the largest connected cluster
   - N_G/N increases once $\langle k\rangle$ exceeds 1 ($\exists$ large clusters) when $\langle k\rangle = 1$, $p \approx \frac{1}{N}$ (sufficient and necessary condition)
     - Subcritical Regime ($0 < \langle k\rangle < 1, p < \frac{1}{N}$) $N_G \sim \ln N$, thus $\lim\limits_{N\to\infty} N_G/N = 0$
     - Critical Point ($1 = \langle k\rangle, p = \frac{1}{N}$) $N_G \sim N^{2/3}$, thus $\lim\limits_{N\to\infty} N_G/N = 0$
     - Supercritical Regime ($1 < \langle k\rangle, p > \frac{1}{N}$) $N_G \sim (p - p_c)N$, thus $\lim\limits_{N\to\infty} N_G/N = 0$
     - Connected Regime($\langle k\rangle > \ln N$, $p > \frac{\ln N}{N}$) $N_G \sim N$, thus $\lim\limits_{N\to\infty} N_G/N = 0$
7. Real Networks are Supercritical
   however they may contain only one large cluster which contradicts the random network theory

8. Small Worlds
   - also known as six degree separation
   - distance between two random nodes in a network is short
   - $\langle k \rangle^d$ nodes at distance d, let $N(d)$ denote the expected number of nodes within distance d, then,

$$N(d) \approx 1 + \langle k \rangle + \langle k \rangle^2 + \cdots + \langle k \rangle^d = \frac{\langle k \rangle^{d+1} - 1}{\langle k \rangle - 1}$$

   - $N(d_{max}) \approx N$, thus when $\langle k \rangle \gg 1$, $\langle k \rangle^{d_{max}} \approx N$, as a result $\langle d_{max} \rangle \approx \frac{\ln N}{\ln \langle k \rangle}$
   - $\langle d \rangle \approx \frac{\ln N}{\ln \langle k \rangle}$ is often a better approximation

9. Clustering Coefficient
   - Let L_i denote the number of links between k_i neigbors, then

$$L_i \sim b(\frac{k_i(k_i-1)}{2}, p)$$

   - thus for a random network $C_i = \frac{2\langle L_i \rangle}{k_i(k_i-1)} = p = \frac{\langle k \rangle}{N} = \langle C_k \rangle$

   However, in real networks
   - $\frac{\langle C_k \rangle}{\langle k \rangle}$ is independent of N
   - C_k is dependent of k

10. Summary: Real Networks are Not Random

THE SCALE FREE PROPERTY

1. Introduction
2. Power Laws and Scale-Free Networks

   • with the log-log plot of the degree distribution being a straight line, $p_k \sim k^{-\gamma}$ where p_k follows a power law distribution and γ being the degree exponent

   • discrete formulization:

   $p_k = Ck^{-\gamma}$, where C is determined by $\sum_{k=1}^{\infty} p_k = 1$

   thus $C = \frac{1}{\sum_{k=1}^{\infty} k^{-\gamma}} = \frac{1}{\zeta(\gamma)}$,and thus $p_k = \frac{k^{-\gamma}}{\zeta(\gamma)}$

   • continuous formulization:

   $p_k = Ck^{-\gamma}$, where C is determined by $\int_{k_{min}}^{\infty} p_k = 1$

   thus $C = \frac{1}{\int_{k_{min}}^{\infty} k^{-\gamma} \mathrm{d}k} = (\gamma - 1)k_{min}^{\gamma-1}$, and thus

   $p(k) = (\gamma - 1)k_{min}^{\gamma-1} k^{-\gamma}$

3. Hubs

• exponential vs power law

- exponential is formed $p(k) = Ce^{-kr}$ where C is bounded by $\int_{k_{min}}^{\infty} p_k \mathrm{d}k = 1$, and k_{max} can be found by $\int_{k_{max}}^{\infty} p_k \mathrm{d}k = \frac{1}{N}$, resulting in $k_{max} = k_{min} + \frac{\ln N}{\lambda}$

- the power law distribution accordingly follows $k_{max} = k_{min} N^{\frac{1}{\gamma - 1}}$

• as a result hubs are fobidden even when N is large for random networks. In scale free networks, hubs are presented naturally

4. The Meaning of Scale Free

• the n^{th} moment of the degree distribution $\langle k^n \rangle = \sum_{k_{min}}^{\infty} k^n p_k \approx \int_{k_{min}}^{\infty} k^n p(k) \mathrm{d}k$

- first moment: average

- second moment: variance

- third moment: skewness

4. (continue)
   - for a scale free network
   $\langle k^n \rangle = \int_{k_{min}}^{k_{max}} k^n p(k) \mathrm{d}k = C \frac{k_{max}^{n-\gamma+1} - k_{min}^{n-\gamma+1}}{n-\gamma+1}$
   if $n - \gamma + 1 \leq 0$, moments that satisfies which is finite
   if $n - \gamma + 1 > 0$, moments that satisfies which is infinite
     - Random Networks have scale
       $\langle k \rangle$ serves as the scale of a random network
       (comparable degree)
     - Scale free network lack scale
5. Universality
     - Plotting Degree Distribution
       - log-log plot, log binding
     - Measuring the Degree Exponent
     - The Shape of p_k for Real Networks

6. Ultra-Small World Property
   - $\langle d \rangle$ scale free is smaller than its random network equivalent
   - the dependent of $\langle d \rangle$ on N and γ can be divided into

$$\langle d \rangle \sim \begin{cases} \text{const.} & \gamma = 2 & \text{Anomalous Regime} \\ \ln \ln N & 2 < \gamma < 3 & \text{Ultra-Small World} \\ \frac{\ln N}{\ln \ln N} & \gamma = 3 & \text{Critical Point} \\ \ln N & \gamma > 3 & \text{Small World} \end{cases}$$

7. The Role of The Degree Exponent
   - three different regimes

   - Anomalous Regime $\gamma \leq 2$
   - Scale-Free Regime $2 < \gamma < 3$
   - Random Network Regime $\gamma > 3$

   the presence of power law follows the equation

$$N = \left(\frac{k_{max}}{k_{min}}\right)^{\gamma-1}$$

8. Generating Networks with Arbitrary Degree Distribution
   - Configuration Model
     - probability to have a link between nodes of degree k_i and k_j follows: $p_{ij} = \frac{k_i k_j}{2L-1}$

     deficiencies: contains self loops and muli-links

8. Generating Networks (Continue)

- Degree-Preserving Randomization
  used to see whether a network property is predicted by its degree distribtuion alone or ir it represents some additional property
  p.s Full Randomization (not degree preserving)
- Hidden Parameter Model use: prevent self-loops and multi-links
  N isolated nodes assigning each node a hidden parameter η_i, where η_i can be from a predefined sequence or a $\rho(\eta)$ distribution. The degree distribution thus follows
  • $p_k = \int \frac{e^{-\eta}\eta_j^k}{k!}\rho(\eta)\mathrm{d}\eta$ if η is from a predefined distribution
  • $p_k = \frac{1}{N}\sum_j \frac{e^{-\eta}\eta_j^k}{k!}$ if η_i is from a deterimistic sequence
  • can create a power law distribution through setting $\eta_j = \frac{c}{i^\alpha}, i = 1, \ldots N$ which makes $p_k \sim k^{-(1+\frac{1}{\alpha})}$. And through this model we can use the fact that $\langle \eta \rangle = \langle k \rangle$ to alter $\langle k \rangle$

9. Summary

- Exponentially Bounded
- Fat Tailed Networks

THE BARABASI-ALBERT MODEL

1. Introduction
2. Growth and Preferential Attachment
   - Growth: Networks expand through the addition of new nodes
   - Preferential Attachment: Nodes prefer to link to more connected nodes
3. The Barabasi-Albert Model

   use: able to generate scale free network

   • start with m_0 nodes

   • network developes in each timestep

   - Growth: each timestep, a new node is added with m links where $m \leq m_0$

   - Preferential Attachment: $\Pi(k_i) = \frac{k_i}{\sum_j k_j}$ is the probability a link of a new node connects to node i with k_i degree

3. The Barabasi-Albert Model(continue)

• the mathematical definition of the Barabasi-Albert model
Linearized Chord Diagram (LCD)
For $m = 1$
(1) start with $G_1^{(0)}$, corresponding to an empty graph with no nodes
(2) Given $G_1^{(t-1)}$ generate G_1^t by adding the node v_t a single link between v_t and v_i where v_i is chosen by

$$p = \begin{cases} \frac{k_i}{2t-1} & \text{if } 1 \leq i \leq t-1 \\ \frac{1}{2t-1} & \text{if } i = t \end{cases}$$

4. Degree Dynamics
   - Assumptions: k_i be continuous
   - $\frac{\mathrm{d}k_i}{\mathrm{d}t} = m\Pi(k_i) = m\frac{k_i}{\sum_{j=1}^{N-1} k_j}$, where $\sum_{j=1}^{N-1} k_j = 2mt - m$,

   thus $\frac{\mathrm{d}k_i}{\mathrm{d}t} = \frac{k_i}{2t-1} \approx \frac{k_i}{2t}$ for large t
   $\rightarrow 2t\mathrm{d}k_i = k_i\mathrm{d}t$ and by intergrating both sides,
   $2tk_i + C = k_i t$ and since $k_i(t_i) = m$
   $C = \ln \frac{t_i^{\frac{1}{2}}}{m}$, thus $k_i(t) = m\left(\frac{t}{t_i}\right)^{\beta}$, where $\beta = \frac{1}{2}$
   thus $\frac{\mathrm{d}k_i(t)}{\mathrm{d}t} = \frac{m}{2}\frac{1}{\sqrt{t_i t}}$

5. Degree Distribution

• continuum theorem:

calculate the degree distribution through

(1) $k_i(t) < k \rightarrow t_i < t\left(\frac{m}{k}\right)^{\frac{1}{\beta}}$

and since $N = m_0 + t$, when t is large $N \approx t$

thus $t\left(\frac{m}{k}\right)^{\frac{1}{\beta}}$ is the fraction of nodes that are more than degree k and $\left(\frac{m}{k}\right)^{\frac{1}{\beta}}$ is the probability for such event

(2) thus $P(k) = 1 - \left(\frac{m}{k}\right)^{\frac{1}{\beta}}$ and take the partial derivative with respect to k, then $p_k \approx 2m^{\frac{1}{\beta}} k^{-\gamma}$ where $\gamma = 3$

• the exact distribution using LCD model is $p_k = \frac{2m(m+1)}{k(k+1)(k+2)}$

6. The Absence of Growth and Preferential Attachment
   - Lack of Preferential Attachment
     $\Pi(k_i) = \frac{1}{(m_0+t-1)}$, and thus $k_i(t) = m \ln \left(e \frac{m_0+t-1}{m_0+t_i-1} \right)$
     as a result $p_k = \frac{e}{m} \exp\left(-\frac{k}{m}\right)$, it follows exponential
   - Lack of Growth
     - adding one link in each timestep without growth
     - $k_i(t) \approx \frac{2}{N} t$ at first is generates a power law, then as $t \to C_2^n$, the graph becomes completely connected

7. Measuring Preferential Attachment
   - two distinct hypothesis preferential attachment relies on:
   - Hypothesis 1: The likelihood to connect to a node dpends on that node's degree k.(random model is k-independent)
   - Hypothesis 2: The functional form of $\Pi(k)$ is linear in k
   - $\frac{\Delta k_i}{\Delta t} \sim \Pi(k_i)$

   to reduce the noise, we use the cumulative preferential attachment function $\pi(k) = \sum_{ki=0}^{k} \Pi(k_i)$
   without preferential attachment, thus $\pi(k) \sim k$. If preferential attachment is linear, $\pi(k) \sim k^2$, $\Pi \sim k^{\alpha}$,
   when $\alpha = 0.9 \pm 0.1$, sublinear is present

8. Non-linear Preferential Attachment
    - Sublinear Preferential Attachment ($0 < \alpha < 1$)
        - $p_k \sim k^{-\alpha} \exp\left(\frac{-2\mu(\alpha)}{\langle k \rangle (1-\alpha)} k^{1-\alpha}\right)$
        - $k_{max} \sim (\ln t)^{1/(1-\alpha)}$
    - Superlinear Preferential Attachment ($\alpha > 1$)
        - $k_{max} \sim t$
9. The Origins of Preferential Attachment (2 kinds of views):
    - Local Mechanism
        - Link Selection Model (create preferential attachment)

        - Growth + Link Selection (both one at a time)

        - $q_k = Ckp_k$, where q_k stands for the probability that the node of a randomly chosen link has degree k

        - given the normalization condition we have $q_k = \frac{kp_k}{\langle k \rangle}$
        - Copying Model

        - Random Connection: probability p the new node links to a random node u

        - Copying: probability $1 - p$ of connecting to an outgoing node of u

        - thus $\Pi(k) = \frac{p}{N} + \frac{1-p}{2L}k$

        (since $\frac{1}{N}$ of selecting a node u and $\frac{k}{2L}$ of selecting a degree-k node through copying)

9. Origins
   - Optimization
     - Cost function: $C_i = \min_j[\delta d_{ij} + h_j]$
     where: d_{ij} is the distance between new node i and a potential target j, h_{ij} is the distance of node j to the first node of the network (serves as the center)
     - Depending on δ:
       - Star Network ($\delta < (1/2)^{(1/2)}$)
         - nodes connect to the center
       - Random Networks ($N^{1/2} \leq \delta$)
         - nodes connect to the closest
       - Scale Free Network ($4 \leq \delta \leq N^{1/2}$)

     Sum: Basin of Attraction (optimization) + Randomness (Random location of new node)

10. Diameter and Clustering Coefficient
    - for the Barabasi Albert network
      - Diameter
        $\langle d \rangle \sim \frac{\ln N}{\ln \ln N}$
      - Clustering Coefficient
        $\langle C \rangle \sim \frac{(\ln N)^2}{N}$

11. Summary

EVOLVING NETWORKS

1. Introduction
   - review: under Barabasi-Albert model, $k(t) \sim t^{\frac{1}{2}}$, which indicates the first mover's advantage
2. The Bianconi-Barabasi Model
   - Property fitness: intrinsic properties of some nodes
     - Growth: a new node j with m links and fitness η is added per timestep, where η is a random number chosen from a fitness distribution $\rho(\eta)$, and it does not change once assigned
     - Preferential Attachment: Probability that a link of a new node connects to node i is $\Pi_i = \frac{\eta_i k_i}{\sum_j \eta_j k_j}$
   - Degree Dynamics

   the degree of node i changes as $\frac{\partial k_i}{\partial t} = m \frac{\eta_i k_i}{\sum_j \eta_j k_j}$

   assuming time evolution of k_i follows a power law with fitness dependent exponent $\beta(\eta_i)$,then $k(t, t_i, \eta_i) = m \left(\frac{t}{t_i}\right)^{\beta(\eta_i)}$

   we get $\beta(\eta) = \frac{\eta}{C}$ with $C = \int \rho(\eta) \frac{\eta}{1-\beta(\eta) d\eta}$

2. The Bianconi-Barabasi Model(continue)

- Degree Distribution is $p_k = C \int d\eta \frac{\rho(\eta)}{\eta} \left(\frac{m}{k}\right)^{\frac{C}{\eta}+1}$
  - Equal Fitnesses
    assuming $\rho(\eta) = \delta(\eta - 1)$, thus $\eta = 1$ and $C = 2$, $\beta = 1/2$, $p_k \sim k^{-3}$, which is the scaling of a Barabasi Albert model
  - Uniform Fitness Distribution
    assuming η comes from $U(0,1)$, thus
    $\exp(-2/C) = 1 - 1/C \to C^{*} \approx 1.255$ $\beta(\eta_i) = \frac{\eta_i}{C^*}$, and thus
    $p_k \sim \int_0^1 d\eta \frac{C^*}{\eta} \frac{1}{k^{(1+C^*)/\eta}} \sim \frac{k^{-(1+C^*)}}{\ln k}$, obtaining $\gamma = 2.255$

3. Measuring Fitness

- reletive difference between degrees $\frac{k_2 - k_1}{k_1} \sim t^{\frac{\eta_2 - \eta_1}{C}}$, thus for large t is quite significant
- take the citation network as an example:

- $\Pi_i \sim \eta_i c_i^t P_i(t)$ where c_i^t is the cumulative number of citations acquired by paper i at time t, $P_i(t)$ indicates that the novelty of citation fades with time where $P_i(t) = \frac{1}{\sqrt{2\pi t}\sigma} e^{-\frac{(\ln t - \mu_i)^2}{2\sigma^2}}$

- fitness is typically exponentially bounded

4. Bose-Einstein Condensation

• Purpose: to see the effect of fitness distribution on network topology

- Fitness $\rightarrow$ Energy

  - assigning fitness with an energy $\varepsilon_i = \frac{1}{\beta_T} \log \eta_i$
- Links $\rightarrow$ Particles
- Nodes $\rightarrow$ Energy Level

• Scale-free Phase

- fit-gets-rich, any moment in this phase, network is scale-free
- uniform fitness distribution is in this phase
- largest hub grows sublinearly

• Bose-Einstein Condensation

- fittest node grabs a finite fraction of links turning into a super hub (winner takes all)
- presence or absense depends on the finess distribution, occur when $\rho(\eta) = (1-\zeta)(1-\eta)^{\zeta}$

5. Evolving Networks

- Review: restriction of Barabasi-Albert Model

- predicts $\gamma = 3$
- ignore small-degree saturation and high degree cutoff
- ignore addition and removal of new links/nodes

- Initial Attractiveness

to allow unconnected nodes to add links, assuming $\Pi(k) \sim A + k$, where A is the initial attractiveness ($\Pi(0) = A$)

- Increase the Degree Exponent $\rightarrow \gamma = 3 + \frac{A}{m}$
- Generate Small Degree Saturation
  $p_k = C(k + A)^{-\gamma}$, which creates less small degree

- Internal Links

- adding new links between prexisting nodes
- $\Pi(k, k') \sim (A + Bk)(A + Bk')$
  - Double Preferential Attachment(A=0)
    consider adding a new node with m links and n internal links in each timestep, then $\gamma = 2 + \frac{m}{m+2n} \rightarrow$ increases heterogenity
  - Random Attachment(B=0)
    - add links between random node pairs, $\gamma = 3 - \frac{2n}{m}$

5. Evolving Networks (Continue)

- Node Deletion

  - each timestep, add a new node with m links and with rate r of single-node removal, r-dependent results are as follows:

    - Scale-free Phase ($r < 1$)
      - removal $<$ new nodes added $\rightarrow$ scale-free with $\gamma = 3 + \frac{2r}{1-r}$
    - Exponential Phase ($r = 1$)
    - Declining Networks ($r > 1$)

- Accelerated Growth

  - degrees increase linearly with time, $L = \langle k \rangle N/2$, where $\langle k \rangle$ is independent of time, yet, degrees may increase faster than N
  - assuming arrival of new links follows $m(t) = m_0 t^{\theta}$
  if $\theta = 0$ m stays is the same
  if $\theta > 0$ the network follows accelerated growth and
  $\gamma = 3 + \frac{2\theta}{1-\theta}$, for $\theta = 1$, leads to hyper-accelerating growth

5. Evolving Networks(Continue)

- Aging

$\Pi(k_i, t - t_i) \sim k(t - t_i)^{-v}$

depending on v, three regimes can be identified

  - $v < 0$

    -network becomes heterogenious
  - $v > 0$

    - new node attach to younger nodes

    - if $v \to \infty$, new nodes contact predecessor immediately
  - $v > 1$

    - effect ovecomes preferential attachment losing the scale free property

6. Summary
   - Topological Diversity
     - Power Law
     - Stretched Exponential (Sublinear Preferential Attachment)
     - Fitness-Induced Corrections
     - Small-Degree Saturation (Initial Attractiveness)
     - High-Degree Cutoffs
   - Modeling Diversity
     - Static Models (fixed nodes and links)
     - Generative Models (configuration, hidden parameter)
     - Evolving Network Models

DEGREE CORRELATION

1. Introduction

   • Paradox: small node link to small nodes, and the probability that degree k node link to k' $P_{kk'} = \frac{kk'}{2L}$ if connection is random

2. Assortativity and Disassortativity

   • three topologies:

     - Neutral Network: follows the equation above
     - Assortative Network: node connect to others with similar degree
     - Disassortative Network: hubs tend to link to small degree nodes

   • network displays degree correlation if links deviate from what is expected by chance

   • degree correlation matrix e_{ij} is the probability of finding nodes of degree i and j at two ends of a randomly selected link and thus $\sum_{i,j} e_{ij} = 1$

2. Assortativity and Disassortativity (Continue)

• $q_k = \frac{kp_k}{\langle k \rangle}$ is the probability that degree k node appears at the end of a randomly selected link, thus $\sum_j e_{ij} = q_i$, and in neutral network $e_{ij} = q_i q_j$

• Disadvantages of Degree using e_{ij} matrix:

  - Difficult to extract information through visual inspection
  - hard to see the magnitude and thus compare networks
  - contains $k_{max}^2/2 \quad r.v.'s$

3. Measuring Degree Correlations

• measure the average degree of its neighbor $k_{nn}(k) = \frac{1}{k_i} \sum_{j=1}^{N} A_{ij} k_j$

• degree correlation function: calculate all node with degree k, $k_{nn} = \sum_{k'} P(k'|k)$ where $P(k'|k)$ is the conditioanal probability that following a link of a k-degree node we reach a degree k' node

3. Measuring Degree Correlations

- Neutral Network

$P(k'|k) = \frac{e_{kk}}{\sum_{k'} e_{kk'}} = \frac{e_{kk'}}{q_k} = \frac{q_{k'} q_k}{q_k} = q_k$

$k_{nn}(k) = \sum_{k'} k' q_{k'} = \sum_{k'} k' \frac{k' p(k' p(k'))}{\langle k \rangle} = \frac{\langle k^2 \rangle}{\langle k \rangle}$

thus results in a friendship paradox (more popular)

- Assortative Network

\- $k_{nn}(k) increases$

- Disassortative Network

\- $k_{nn}(k) decreases$

- by empirical results, we estimate $k_{nn}(k) = ak^{\mu}$ where:
  - Assortative Network ($\mu > 0$)
  - Neutral Network ($\mu = 0$)
  - Disassortative Network ($\mu < 0$)

- a single number to capture magnitude of correlation, either

by μ or correlation coefficient $r = \sum_{jk} \frac{jk(e_{jk} - q_j q_k)}{\sigma^2}$

where $\sigma^2 = \sum_k k^2 q_k - [\sum_k k q_k]^2$

$r > 0$ assortative, $r < 0$ disassortative, and $r = 0$ neutral

4. Structural Cutoffs

• Conflict: the expected number of links between degree k and k' node is $E_{kk'} = e_{kk'} \langle k \rangle N$, in a neutral network, $E_{kk'} = \frac{k p_k k' p_{k'}}{\langle k \rangle} N > 1$
• only for nodes that exceed some k_s does the conflict happens, known as structural cutfoff, where $k_s \sim (\langle k \rangle N)^{1/2}$
• comparing k_{max} and k_s, in scale free: $k_{max} \sim N^{\frac{1}{\gamma - 1}}$

- No Structural Cutoff ($k_{max} < k_s$)
- Structural Disassortivity ($k_{max} > k_s$)
  - fewer links between hubs than expected -to generate networks free of structural disassortivity:
    - allow multiple links between nodes (not simple)
    - remove hubs with degree larger than k_s (simple network)

4. Structural Cutoffs (Continue)

• Correlation Due to Structural Disassortativity or Unknown Process

- use the following techniques to test

  - Degree Preserving Randomization with Simple Links (R-S)
    - if $k_{nn}(k)$ and $k_{nn}^{R-S}(k)$ are indistinguishable, then structural
    - if one shows degree correlation and the other doesn't, then there is an unknown process
  - Degree Preserving Randomization with Multiple Links (R-M)
    - always create neutral network
    - compare $k_{nn}(k)$ and $k_{nn}^{R-M}(k)$ to see if

5. Generating Correlated Networks

- Degree Correlation in Static Models
  - Erdos-Renyi Model (Neutral)
  - Configuration Model (Neutral)

    - both follows $k_{nn}(k) = \frac{\langle k^2 \rangle}{\langle k \rangle}$
  - Hidden Parameter Model

    - neutral if multi-links allowed, if not allowed $k_{nn}(k) \sim k^{-1}$, thus being disassortive. (i.e. $\mu = 1$)
- Degree Correlation in Evolving Networks
  - Initial Attractiveness Model ($\Pi(k) \sim A + k$)

    - three regimes
    - Disassortative Regime ($\gamma < 3$)

      - $k_{nn}(k) \sim k^{-\frac{|A|}{m}}$
    - Neutral Regime ($\gamma = 3$) - $k_{nn}(k) \sim \frac{m}{2} \ln N$
    - Weak Assortativity Regime ($\gamma > 3$)

      - $k_{nn}(k) \approx (m + A) \ln \left(\frac{k}{m+A} \right)$
  - Bianconi-Barabasi Model

    - uniform fitness: disassortive network (structural), however does not fully overlap, meaning disassortivity not fully explained by scale free

5. Generating Correlated Networks (Continue)
   - Tuning Degree Correlations

   - purpose: generate network with predefined degree distribution with max correlation

Step 1 Link Selection:
- choose at random two links, label four nodes at the end of the two links and order them in accordance to the degree value

Step 2 Rewiring

2A Assortative:
- pair two highest node and two lowest nodes

2B Disassortative:
- pair the highest and lowest nodes

if simple network is required, check if rewiring leads to multilinks, and if it does reject and redo step 1

6. The Impact of Degree Correlations
   - difference in transition phase of giant components
     - Assortative Network
       - giant component emerges even when $\langle k \rangle < 1$
     - Disassortative Network
       - difficult forming giant component
     - Giant Component
       - for large k giant component is smaller for assortative since its harder to attract small degree nodes
   - additional consequences:
     - for assortative network, $\langle d \rangle$ is smaller, however, $\langle d_{max} \rangle$ is larger (the later is due to increase in $k = 2$ nodes)
     - stability of the network
7. Summary
   - Complete Degree Correlation is determined by $P(k^{(1)}, k^{(2)}, \ldots, k^{(k)} | k)$, a node with degree k connects to node with degree $k^{(1)}, k^{(2)}, \ldots, k^{(k)}$

NETWORK ROBUSTNESS

1. Introduction
2. Percolation Theory

• Problem: given a lattice, if each intersection has p of being placed a pebble, what is ...

- the expected size of the largest cluster?
- the size of the average cluster?

$\rightarrow$ the larger the p the larger the cluster, thus assuming p_c to be threshold of percolating cluster, we define:

- Average Cluster Size ($\langle s \rangle$)
  - $\langle s \rangle \sim |p - p_c|^{-\gamma_p}$, diverges as $p \rightarrow p_c$
- Order Parameter (P_∞)
  - the probability that a randomly chosen pebble belongs to the largest cluster $P_\infty \sim (p - p_c)^{\beta_p}$, 0 as $p \rightarrow p_c$
- Correlation Length (ζ)
  - the mean distance between two pebble that belong to the same cluster $\zeta \sim |p - p_c|^{-v}$

- critical exponents: γ_p, β_p, v, are universal, i.e. independent of the lattice nature or p_c

2. Percolation Theory (Continue)
   - p_c depends on lattice type
   - p_c depends on lattice dimension (for large dimension, independent)
   - critical exponents depend on lattice dimension alone

   • Inverse Percolation Transition and robustness

   - randomly remove f fraction of nodes, for f larger than f_c giant component vanishes.

   - P_∞ is dependent on f and f_c

   - $f = 1 - p$

3. Robustnesss of Scale-Free Networks

   • percolation focus on graphs with nodes of identical degree

   • high f_c for scale free network

   • Molloy-Reed Criterion

   - for arbitrary distribution, for giant component to exist, connect to at least 2 nodes on average k$= \frac{\langle k^2 \rangle}{\langle k \rangle} > 2$

3. Robustness of Scale-Free Networks (Continue)
   - Critical Threshold

   - using Molloy-Reed, f_c for arbitrary distribution follows
   $f_c = 1 - \frac{1}{\frac{\langle k^2 \rangle}{k} - 1}$
   - the scale free network's f_c depends on γ with 3 as threshold
   - Robustness of Finite Networks

   - $f_c \approx 1 - \frac{C}{N^{\frac{3-\gamma}{\gamma-1}}}$ where C is a N-independent term
   - perceive relative size of a giant component using $\frac{P_\infty(f)}{P_\infty(0)}$, where f is the fraction removed
4. Attack Tolerance
   - Attack: start with removal of highest degree nodes
   - Critical Threshold Under Attack

   - attack consequences $k_{max} \rightarrow k'_{max}$, $p_k \rightarrow p'_k$
   - scale-free is way more fragile under attacks than random networks, f_c decreases drastically

5. Cascading Failures
   - initial failure keeps spreading
   - $p(s) \sim s^{-\alpha}$, where α is the avalanche exponent
6. Modeling Cascading Failure
   - Failure Propagation Model

   - consider network with arbitrary distribution, each node contains an agent. agent i can be in state 0(healthy) or 1(failed), φ be the breakdown threshold for all φ_i
   - all agents are healthy in $t = 0$, in each timestep, randomly pick an agent to update its state
     - if agent i is 0, if at least φ fraction of its k_i neighbors are in state 1, change it to 1
     - if agent i is 1, remain unchanged
   - Three Regimes:
     - Subcritical Regime ($\langle k \rangle$ large)
       - size of cascading event s follows exponential
     - Supercritical Regime ($\langle k \rangle$ small)
     - Critical Regime (boundary of the former two)

7. Modeling Cascading Failure

- Braching Model

- failure propogation model being too complicated
- cascading failure follows a branching process
- node with initial failure is the root of the tree
- Details: starts with a single active node, and produces k offsprings based on the a p_k distribution, if node selects $k = 0$, the branch dies out, if $k > 0$, k new active sites is created
- three regime:
  - Subcritical Regime ($\langle k \rangle < 1$)
    - terminated
  - Supercritical Regime ($\langle k \rangle > 1$)
    - grow indefinitely
  - Critical Regime ($\langle k \rangle = 1$)
    size s follows a power law
- Use: branching model can help determine the avalanche size distribution $p(s)$ through p_k

8. Building Robustness
   - Design Robust Network

   - Robustness Captured by f_c which depends on $\langle k^2 \rangle$ and $\langle k \rangle$, thus the goal is to maximize $\langle k^2 \rangle$ and keep $\langle k \rangle$ (cost) fixed

   - take into account random failures and attack, thus goal is to maximize $f_c^{tot} = f_c^{rand} + f_c^{targ}$ - best achieved by $p_k = (1 - r)\delta(k - k_{min}) + r\delta(k - k_{max})$, r is frac of node with degree k_{max}, thus, best achieved when $r = 1/N$
9. Summary