Dynamic Network Properties and Models

Prof. Shou-de Lin
CSIE/GINM, NTU


Overview

• Review: network generating models
• Two properties in the view of Conventional wisdom vs. New findings, and empirical studies
  – Densification Power Law
  – Shrinking diameter
• Two Dynamic Network Generation Models
  – Community Guided Attachment (CGA)
  – Fire Forest Model (FF)
• Phase Plots (Sharp phase transition)
• Densification and degree distribution over time
Network Generating Models

• A network model is a **process** (randomized or deterministic) for **generating a network to satisfy** certain patterns or phenomenon as observed

• **Static** network Models
  – Care about the static property of a social network,
    • such as the average path length, clustering coefficient.

• **Dynamic** networks models
  – Care about the dynamic property of a social network
    • such as how the average path length changes when new nodes and edges are added
# Review: Static Network Models & Static Properties

<table>
<thead>
<tr>
<th></th>
<th>ER Model</th>
<th>Regular Lattice</th>
<th>WS Model</th>
<th>BA Model</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Average Path Length</strong></td>
<td>Short</td>
<td>Long</td>
<td>Short</td>
<td>Short</td>
</tr>
<tr>
<td></td>
<td>$l_{\text{rand}} \sim \log(N)/\log(z)$</td>
<td>$l_{\text{lattice}} \sim N/2K$</td>
<td>$l_{\text{WS}} \sim \log(N)$</td>
<td>$l_{\text{sf}} \sim \log(N)$</td>
</tr>
<tr>
<td><strong>Clustering Coefficient</strong></td>
<td>Low</td>
<td>High</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>$C_{\text{rand}} = z/n$</td>
<td>$C_{\text{lattice}} = 3/4$</td>
<td>$C_{\text{WS}}(p) = C(0)(1-p)^3$ independent of $N$</td>
<td>$C_{\text{sf}} = N^{-0.75}$</td>
</tr>
<tr>
<td><strong>Long Tail</strong></td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>$P(k) = z^k e^{-z}/k!$</td>
<td>$P(k) = k$</td>
<td>$P(k) = Ck^{-\alpha}$</td>
<td>$P(k) = \alpha k^\alpha$</td>
</tr>
</tbody>
</table>

2009/12/15

SNA09F, Prof. Shou-de Lin
What to Convey in This Lecture

• Explicitly:
  – Several dynamic properties about a dynamic social network (observed from data)
  – Some dynamic network generation models to explain such models

• Implicitly:
  – One case study on how research can be done in this area.
  – It is not necessary the only way or ideal way of doing SNA research, but is apparently a popular way
Why Modeling Dynamic Networks (1/2)

• Accurate properties of network growth, together with models supporting them, have implications in several contexts

1. **Graph generation**
   – Allow assessing the quality of graph generators
   – Synthetic graphs are needed when real graphs are hard to collect

2. **Prediction**
   – Given several snapshots of the past, what can we say about the future of a social network?
Why Modeling Dynamic Networks (2/2)

3. **Graph Sampling**
   - Existing measures (e.g. APL, centralities, PageRank) are impractical for huge networks → need smaller ones
   - Sampling provides subgraphs with the similar properties as the original one

4. **Abnormality detection**
   - In many network settings, normal behavior will produce graphs that obey certain network properties
   - Activity producing structures that deviate significantly from the normal behavior could indicate something interesting
Evolution of a Social Network: Conventional Wisdom

• **Constant average degree assumption**
  – The average node degree in the network remain constant over time
  – Or equivalently, the number of edges grows linearly in the number of nodes

• **Slowly growing diameter assumption**
  – The diameter is a slowly growing function of the network size (e.g. logN)
Evolution of a Social Network: New Empirical Findings

1. **Densification Power laws**
   - The networks are becoming denser over time, with average degree increasing
   - The number of edges growing super-linearly in the number of nodes
   - The densification follows power-law

2. **Shrinking Diameters**
   - The effective diameter is, in many cases, actually decreasing as the network grows

- 1 and 2 are intuitively consistent
  - But it is possible to create a network that has one property but not the other.
Densification Power Law (1/3)

• N(t) : nodes at time t
• E(t) : edges at time t
• Suppose that
  \[ N(t+1) = 2 \times N(t) \]
• Q: what is your guess for \( E(t+1) \)?
  \[ E(t+1) \neq 2 \times E(t) \]
• Ans: more-than 2 times (something like \( 2^k \))!
  – obeying the Densification Power Law
Densification Power Law (2/3)

- Densification Power Law
  - Networks are becoming denser when grows
  - The number of edges grows faster than the number of nodes – average degree is increasing

\[ E(t) \propto N(t)^a \]

or equivalently

\[ \frac{\log(E(t))}{\log(N(t))} = \text{const} \]

\( a \) : densification exponent
Densification Power Law \((3/3)\)

\[
E(t) \propto N(t)^a
\]

- Densification exponent: \(1 \leq a \leq 2\)
  - \(a=1\): constant degree (assumed in the literature so far)
  - \(a=2\): almost fully connected graph— on average, edges to a constant fraction of all nodes
Real Case 1: Physics Citations

- Citations among physics papers
- 1992:
  - 1,293 papers, 2,717 citations
- 2003:
  - 29,555 papers, 352,807 citations
- For each month $M$, create a graph of all citations up to month $M$
Real Case 2: Patent Citations

- Citations among patents granted
- 1975
  - 334,000 nodes
  - 676,000 edges
- 1999
  - 2.9 million nodes
  - 16.5 million edges
- Each year is a datapoint

\[ E(t) \]

\[ N(t) \]

\[ = 0.0002 \times 1.66 \]

\[ R^2 = 0.99 \]
Real Case 3: **Autonomous Systems**

- **Graph of Internet**
  - 1997
    - 3,000 nodes
    - 10,000 edges
  - 2000
    - 6,000 nodes
    - 26,000 edges
- **One graph per day**
Real Case 4: Affiliation Network

- Authors linked to their publications
- 1992
  - 318 nodes
  - 272 edges
- 2002
  - 60,000 nodes
    - 20,000 authors
    - 38,000 papers
  - 133,000 edges
Densification Power Law on More Real-world Networks

Table I. Dataset Names with Sizes, Time Lengths and Densification Power Law Exponents

Notice the very high densification exponent for citation networks ($\approx 1.6$), around $1.2$ for Autonomous Systems, and lower (but still significant) densification exponent ($\approx 1.1$) for affiliation and collaboration type networks.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Nodes</th>
<th>Edges</th>
<th>Time</th>
<th>DPL exponent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arxiv HEP–PH</td>
<td>30,501</td>
<td>347,268</td>
<td>124 months</td>
<td>1.56</td>
</tr>
<tr>
<td>Arxiv HEP–TH</td>
<td>29,555</td>
<td>352,807</td>
<td>124 months</td>
<td>1.68</td>
</tr>
<tr>
<td>Patents</td>
<td>3,923,922</td>
<td>16,522,438</td>
<td>37 years</td>
<td>1.66</td>
</tr>
<tr>
<td>AS</td>
<td>6,474</td>
<td>26,467</td>
<td>785 days</td>
<td>1.18</td>
</tr>
<tr>
<td>Affiliation ASTRO–PH</td>
<td>57,381</td>
<td>133,179</td>
<td>10 years</td>
<td>1.15</td>
</tr>
<tr>
<td>Affiliation COND–MAT</td>
<td>62,085</td>
<td>108,182</td>
<td>10 years</td>
<td>1.10</td>
</tr>
<tr>
<td>Affiliation GR-QC</td>
<td>19,309</td>
<td>26,169</td>
<td>10 years</td>
<td>1.08</td>
</tr>
<tr>
<td>Affiliation HEP–PH</td>
<td>51,037</td>
<td>89,163</td>
<td>10 years</td>
<td>1.08</td>
</tr>
<tr>
<td>Affiliation HEP–TH</td>
<td>45,280</td>
<td>68,695</td>
<td>10 years</td>
<td>1.08</td>
</tr>
<tr>
<td>Email</td>
<td>35,756</td>
<td>123,254</td>
<td>18 months</td>
<td>1.12</td>
</tr>
<tr>
<td>IMDB</td>
<td>1,230,276</td>
<td>3,790,667</td>
<td>114 years</td>
<td>1.11</td>
</tr>
<tr>
<td>Recommendations</td>
<td>3,943,084</td>
<td>15,656,121</td>
<td>710 days</td>
<td>1.26</td>
</tr>
</tbody>
</table>
Shrinking Diameter

• Prior work on Power Law graphs hints at **Slowly growing diameter**
  – diameter $\sim O(\log N)$
  – diameter $\sim O(\log \log N)$

• However, some observation on the real data shows that **Diameter shrinks over time**
  – As the network grows the distances between nodes slowly decrease
Real Cases: ArXiv citation graph & Autonomous Systems

- Citations among physics papers
- One graph per year

- Graph of Internet
- One graph per day
- 1997 – 2000
Real Cases: Affiliation Network & Patent Network

- Graph of collaborations in physics – authors linked to papers
- 10 years of data

- Patent citation network
- 25 years of data
Validating Shrinking Diameter

• Some factors could affect shrinking diameter
  – Possible Sampling Problem
    • Some approximating methods are applied to compute effective diameter
  – Disconnected components
  – “Missing Past” Effect
    • How do we handle the citations outside the dataset?
  – Emergence of Giant Component

• None of them matters
Two Dynamic Network Models

• Existing graph generation models do NOT capture the **Densification Power Law** and **Shrinking diameter**
  – Can we find a **simple model of local behavior**, which naturally leads to the observed phenomena?

• There are two models:
  – **Community Guided Attachment (CGA)**
    • Obey Densification Power Law
  – **Forest Fire Model (FF)**
    • Obey Densification Power Law
    • Follow shrinking diameter (and Power Law degree distribution)
Community Guided Attachment: Motivation

• What are the underlying principles that drive a social network to obey a densification power law, without central control or coordination?

• Consider a brute-force model
  – Define a graph model in which \( E(t) \sim N(t)^a \) by simply having each node, when it arrives at time \( t \), generate \( N(t)^{a-1} \) outlinks
  – Intuition: each author of a new paper has a rule “cite \( n^{a-1} \) other papers,” hard-wired in her brain
  – However, this does not provide any insight into the origin of the exponent \( a \), as the exponent is unrelated to the operational details of network construction
Community Guided Attachment: Observation

• **Power laws** often appear in combination with **self-similar, recursive** structures
  
  – Ex1. Computer networks form tight groups (e.g. based on geography), which consist of smaller groups recursively
  – Ex2. Patents form conceptual groups (e.g. chemistry, communication), which consist of sub-groups

• **“Community within communities patterns”**
  
  – 1. Social structures exhibit self-similarity, with individuals organizing their **social hierarchy**
  – 2. Pairs of individuals belonging to the same small community **form social ties more easily** than pairs of individuals who are only related by membership in a **larger community**
Community within Communities Patterns

• One expects many within-group friendships and fewer cross-group ones
• How hard is it to cross communities?

Self-similar university community structure
Basic Version of CGA

• Represent the recursive structure of communities-within-communities as a tree $\Gamma$, of height $H$. Here we will show that a simple, perfectly balanced tree of constant fanout $b$ is enough to lead to a densification power law.

• Nodes $V$ in the graph are represented as the leaves of the tree
  – $n=|V|$, $n=b^H$
  – Let $h(v,w)$ be the tree distance of two leaf nodes $v$ and $w$ (i.e., the height of their least common ancestor)
Basic Version of CGA

• A random graph on a set of nodes V by specifying the probability that v and w form a link as a function \( f \) of \( h(v,w) \)
  – \( f \): difficulty function, it should decrease with \( h \)

• It’s better to have a scale-free \( f \), that is, \( f(h)/f(h-1) = \text{constant} \). Therefore

\[
f(h) = c^{-h}
\]

\( c \) is the difficulty constant
\( c \geq 1 \)

Cross-communities links become harder to form as \( c \) increases
CGA Properties

• In the community Guide Attachment (CGA) random graph model, the expected average out-degree \( d \) of a node is proportional to

\[
\bar{d} = n^{1-\log_b(c)} \quad \text{if } 1 \leq c \leq b \\
= \log_b(n) \quad \text{if } c = b \\
= \text{constant} \quad \text{if } c > b
\]
Proof

• For each given node $v$, the expected out-degree of the node is proportional to

$$d = \sum_{x \neq v} f(h(x, v)) = \sum_{j=1}^{\log_b(n)} (b - 1)b^{j-1}c^{-j} = \frac{b-1}{c} \sum_{j=1}^{\log_b(n)} \left(\frac{b}{c}\right)^{j-1}$$

• There are three different cases:

  – If $1 \leq c < b$, then by summing the geometric series

    $$d = \frac{b-1}{c} \cdot \left(\frac{b}{c}\right)^{\log_b(n)} - 1 = \left(\frac{b-1}{b-c}\right) \left(n^{1-\log_b(c)} - 1\right) = \Theta(n^{1-\log_b(c)})$$

  – If $c = b$, the series sums to

    $$d = \sum_{x \neq v} f(h(x, v)) = \frac{b-1}{b} \sum_{j=1}^{\log_b(n)} \left(\frac{b}{b}\right)^{j-1} = \frac{b-1}{b} \log_b(n) = \Theta(\log_b(n))$$

  – If $c > b$, it converges to a constant
Analysis of CGA

• Note that when $c < b$, we get a densification law with exponent $\alpha$ greater than 1
  
  – The **expected out-degree** is $n^{1 - \log_b(c)}$
  
  – So the **total number of edges** grows as $n^\alpha$

where

$$\alpha = 2 - \log_b(c)$$

– Moreover, as $c$ varies over the interval $[1,b)$, the exponent $\alpha$ ranges over all values in the interval $(1,2]$
Difficulty Constant

\[ a = 2 - \log_b(c) \]

• Given different non-integer densification exponent
  – If \( c = 1 \): easy to cross communities then: \( a = 2 \), **quadratic growth** of edges, near clique
  – If \( c = 2 \): hard to cross communities then: \( a = 1 \), **linear growth** of edges, constant out-degree
Densification Power Law

• The **Community Guided Attachment** leads to **Densification Power Law** with exponent

\[ a = 2 - \log_b(c) \]

– a: densification exponent \( E(t) \propto N(t)^a \)
– b: community structure branching factor
– c: difficulty constant
Dynamic Community Guide Attachment
(Dynamic CGA)

• In CGA
  – Nodes are first organized into a nested set of communities
  – Then they start forming links

• In the Dynamic CGA model
  – Nodes are added over time and the nested structure deepens to accommodate them

• Assumption
  – A node **only creates out-links** at the moment it is added (i.e., **only link to nodes already present**)
  – e.g. a paper usually cites those before itself
Dynamic CGA

• Rather than having graph nodes reside only at the leaves of the tree $\Gamma$, now it allows a graph node corresponding to every internal node in $\Gamma$

• Initially (Step-1)
  – There is a single node $v$ in the graph, and the tree $G$ consists just of $v$

• In time step $t$ (Step-2)
  – Go from a complete b-ary tree of depth $t-1$ to one of depth $t$, by adding $b$ new leaves as children of each current leaf
  – Each of these new leaves will contain a new node of the graph
Dynamic CGA (cont.)

- In time step t (Step-3)
  - Each new node forms out-links as follows
  - For a constant $c$, node $v$ forms a link to each node $w$, independently, with probability $\frac{c - d(v,w)}{2}$

**Explanation:**

- Since a new node has the ability to link to internal nodes of the existing tree, not just to other leaves
- Here we define the **tree-distance** $d(v,w)$ between nodes $v$ and $w$ to be
  - $|\text{path from } v \text{ up to the least common ancestor of } v\&w| + |\text{path from the least common ancestor down to } w|$  
  - If $v$ and $w$ are both leaves, $d(v,w)=2h(v,w)$
For a constant $c$, node $v$ forms a link to each node $w$, independently, with probability

$$p = c^{-d(v,w)/2} = c^{-3/2}$$

$$p = c^{-d(h,i)/2} = c^{-2/2}$$

$$p = c^{-d(i,k)/2} = c^{-4/2}$$
Properties of Dynamic CGA

• When $c < b$ (densification power law)
  – Average node degree = $n^{1-\log_b(c)}$

• Proof:
  – The leave nodes have higher average degree than other nodes.
  – The expected out-degree for a leave node is bounded as
    $b^1c^{-1}+b^2c^{-2}+\ldots b^\log n c^{-\log n} = O((b/c)^{\log n})=O(n^{1-\log c})$
  – Since a constant fraction of all nodes are leaves, the expected out-degree for all nodes are bounded to
    $O(n^{1-\log c})$ as well

• When $b < c$
  – Average node degree = constant, since
    $O((b/c)^{\log n})=O(1)$

• Recall: a node only link to nodes already present
  a node only links to nodes whose height is higher than or equal to it
Fire Forest Model: Motivation & Goal

• Disadvantage for CGA
  – Require explicit community structure
  – Cannot explain for shrinking diameter

• Fire Forest Model
  – Given a (possibly empty) initial graph $G$, a sequence of new nodes $v_1, \ldots, v_k$
  – A simple randomized process to successively link $v_i$ to nodes of $G$ ($i=1,\ldots,k$)
  – So that the resulting graph $G_{final}$ will obey:
    1. Heavy-tail in-degree and out-degree distribution (needs some flavor of rich-get-richer)
    2. Emergence of communities (without assuming community hierarchy)
    3. Densification Power Law (need some flavor of community guided attachment)
    4. Shrinking Diameter
Forest Fire Model: Intuition

- **How do authors identify references?**
  1. Find a first paper and cite it
  2. Chase and cite a subset of the references in the paper (modeled here as random)
  3. Continues recursively with the paper discovered in this way
  4. Depending on the bibliographic tools (e.g. CiteSeer) being used in this process, it may also be possible to chase back-links to papers that cite the paper under consideration
FF Algorithmic Process

• Two parameters
  – Forward Burning Probability $p$
  – Backward Burning ratio $r$

• Consider a node $v$ joining the network at time $t > 1$, and let $G_t$ be the graph constructed thus far ($G_1$ will consist of just a single node).

• Node $v$ forms out-links to nodes in $G_t$ according to the following process
FF Algorithmic Process

1. v first chooses an **ambassador node w** uniformly at random, and forms a link to w

2. (1) Generate two random numbers, x and y, following **geometric distribution** with means \( p/(1-p) \) and \( rp/(1-rp) \).
   (2) Randomly select x out-neighbors of w as \( w_1...w_x \) and y in-neighbors of w as \( w_{x+1}...w_{x+y} \)

3. v forms out-links to \( w_1, w_2, ..., w_{x+y} \), and then applies step-2 recursively to each of \( w_1, w_2, ..., w_{x+y} \). Nodes cannot be visited a second time.
Forest Fire Model: Example

• A node arrives
• Randomly chooses an “ambassador”
• Starts burning nodes (with probability $p$) and adds links to burned nodes
• “Fire” spreads recursively
Why FF Model Captures Several Network Properties

• **Heavy-tailed in-degree:** because FF model has a rich-get-richer flavor: highly-linked nodes have higher chance to be reached by newcomers.

• **Communities:** a newcomer copies several of the neighbors of his/her ambassador, therefore can form community easily. [Stochastic models for the web graph, Kumar et al.2000]

• **Densification power law:** a newcomer will have more links near the community of the ambassador, a few links beyond this, and significantly fewer farther away. Analogous to the CGA.
Validation via Simulations

• By varying just two parameter $p$ and $r$, it is possible to produce graphs that densify ($a>1$), exhibit heavy-tailed distributions for both in- and out-degrees, and have shrinking diameters.

E(t) densification

Number of edges

10^1 10^2 10^3 10^4 10^5

Number of nodes

10^1 10^2 10^3 10^4 10^5

2009/12/15

* Edges

= 0.83 x^{1.21} R^2 = 1.00

2009/12/15

Number of nodes

Log fit

0 2000 4000 6000 8000 10000

Number of nodes

5 5.5 6 6.5 7 7.5 8 8.5

SNA09F, Prof. Shou-de Lin
Validation via Simulations

- **in-degree**
  - count vs. in-degree
  - Log-log scale

- **out-degree**
  - count vs. out-degree
  - Log-log scale

2009/12/15
SNA09F, Prof. Shou-de Lin
Densification and Shrinking Diameter

- Q: Are Densification and Shrinking Diameter two different observations of the same phenomena?
- A: No!
- Forest Fire can generate:
  1. **Sparse** graphs with increasing diameter
     \( a=1.01<2, \ p=0.35, \ p_b=0.2 \)
  2. **Sparse** graphs with decreasing diameter
     \( a=1.21<2, \ p=0.37, \ p_b=0.32 \)
  3. **Dense** graphs with decreasing diameter
     \( a\sim2, \ p=0.38, \ p_b=0.35 \)
Natural Extensions to FF Model

• Orphans
  – In many real-world networks, there are many isolated nodes
    • e.g. papers with no citations to certain corpus
  – [Adjusted Model] Use another probability $q>0$ to model a newcomer that forms no links
  – This will have a more pronounced decrease in diameter over time, with large distances caused by groups of nodes linking to different orphans gradually diminishing as further nodes arrive to connect them together
Natural Extensions to FF Model (cont.)

• **Multiple ambassadors**
  
  – Allow newcomers to choose more than one ambassador with some positive probability
  
  – i.e., rather than burning links starting from just one node, a newly arriving node burns links starting from two or more
  
  – This extension accentuates the decrease in diameter over time, as nodes linking to multiple ambassadors serve to bring together formerly far-apart parts of the graph
Phase Plot (1/4)

• Here explore the parameter space of an FF model in terms of forward-burning prob. $p$ and backward-burning ratio $r$ (backward-burning prob. $p_b=rp$)

• Effective diameter log-fit factor $\alpha$
  
  – The logarithmic function of the form
  \[
  \text{diameter} = \alpha \log t + \beta 
  \]
  
  – If $\alpha < 0$: decreasing effective diameter over time
  – If $\alpha > 0$: increasing effective diameter
Phase Plot (2/4)

- **Sharp transition** of densification power law
- At low \( p \), increasing diameter & no densification (\( \alpha=1 \))
- As \( p \) increases, diameter grows slower and slower
- For the narrow band, decreasing diameter & negative \( \alpha \)
- With high \( p \), diameter is constant (\( \alpha \sim 0 \), diameter \( \sim 1 \), a \( \sim 2 \)): clique

\[ E(t) \propto N(t) \alpha \]
\[ \text{diameter} = \alpha \log t + \beta \]

(a) We fix burning ratio, \( r = 0.5 \) and vary forward-burning probability \( p \)

(b) We fix backward-burning probability \( p_b = 0.3 \) and vary forward-burning probability \( p \)
Phase Plot (3/4)

- Contour plot of densification exponent \( a \)
- Contours in between correspond to a 0.1 increase in DPL exponent
  - The left-most contour corresponds to \( a=1.1 \)
  - The right-most contour corresponds to \( a=1.9 \)
- The desirable region is in between; it is very narrow
  - \( a \) increases dramatically along a contour line, suggesting a sharp transition

The graph is dense, e.g. clique

- Constant average degree
Phase Plot (4/4)  \[ \text{diameter} = \alpha \log t + \beta \]

- Contour plot for the effective diameter log-fit factor \( \alpha \)
- Each contour corresponds to diameter factor \( \alpha \)
- Varying \( \alpha \) in range \(-0.3 \leq \alpha \leq 0.1\), with step-size 0.05
- It is very **narrow in sharp transition** as well
Densification and Degree Distribution over Time

• Here we analyze the relation between the densification and the power-law degree distribution over time
• The analysis can be divided into two cases

1. If the degree distribution of a time evolving graph is power-law
   – It maintains constant power-law exponent $\gamma$ over time
   – We show that for $1 < \gamma < 2$, densification power law with exponent
     \[ a = \frac{2}{\gamma} \]
   – In this case, the densification power law is the consequence of the fact that a power-law distribution

Recall:
\[ P(k) \sim k^{-\gamma} \]
(2 < $\gamma$ < 3)
Densification and Degree Distribution over Time

2. If a temporally evolving graph densifies with densification exponent $a$

- It follows a power-law degree distribution with $\gamma > 2$ that can change over time
- We show that for a given densification exponent $a$, the power-law degree exponent $\gamma_n$ has to evolve with the size of the graph $n$ as

$$\gamma_n = \frac{4n^{a-1} - 1}{2n^{a-1} - 1}$$

- This shows that the densification power law and the degree distribution are related and that one implies the other
Case 1
Constant Degree Exponent over Time

• Assuming the exponent $\gamma$ (slope) of the degree distribution does not change over time, a natural question to ask is:
  
  – Q: What is the relation between densification power law exponent and the degree distribution over time?

• In a temporally evolving graph with a power-law degree distribution having constant degree exponent $\gamma$ over time, the densification power law $a$ is:

$$a = \begin{cases} 
1 & \text{if } \gamma > 2 \\
\frac{2}{\gamma} & \text{if } 1 \leq \gamma \leq 2 \\
2 & \text{if } \gamma < 1 
\end{cases}$$
Proof: Constant Degree Exponent over Time

- Assume at any time $t$, the degree distribution of an undirected graph $G$ follows a power law
  - The number of nodes $D_d$ with degree $d$ is $D_d = cd^{-\gamma}$
  - Assume at some point in time the maximum degree in the graph is $d_{\text{max}}$
  - Later, as the graph grows, we will let $d_{\text{max}} \rightarrow \infty$

- Here we calculate the number of nodes $n$ and the number of edges $e$ in the graph

$$n = \sum_{d=1}^{d_{\text{max}}} cd^{-\gamma} \approx \int_{d=1}^{d_{\text{max}}} d^{-\gamma} = c \frac{d_{\text{max}}^{1-\gamma} - 1}{1 - \gamma}$$

$$e = \frac{1}{2} \sum_{d=1}^{d_{\text{max}}} cd^{1-\gamma} \approx \int_{d=1}^{d_{\text{max}}} d^{1-\gamma} = c \frac{d_{\text{max}}^{2-\gamma} - 1}{2 - \gamma}$$
Proof: Constant Degree Exponent over Time

• Because \( c d_{\text{max}}^{-\lambda} = 1 \), \( \log(c) = \gamma \log(d_{\text{max}}) \)

• Now let the graph grow, so \( d_{\text{max}} \to \infty \)

Then the densification power law exponent \( a \) is

\[
a = \lim_{d_{\text{max}} \to \infty} \frac{\log(e)}{\log(n)} = \frac{\gamma \log(d_{\text{max}}) + \log(|d_{\text{max}}^{2-\gamma} - 1|) - \log(|2 - \gamma|)}{\gamma \log(d_{\text{max}}) + \log(|d_{\text{max}}^{1-\gamma} - 1|) - \log(|1 - \gamma|)}
\]

• Case 1. \( \gamma > 2 \): No densification. \( a = \frac{\gamma \log(d_{\text{max}}) + o(1)}{\gamma \log(d_{\text{max}}) + o(1)} = 1 \)

• Case 2. \( 1 < \gamma < 2 \): Densification arises.

\[
a = \frac{\gamma \log(d_{\text{max}}) + (2 - \gamma)\log(d_{\text{max}}) + o(1)}{\gamma \log(d_{\text{max}}) + o(1)} = \frac{2}{\gamma}
\]

• Case 3. \( \gamma < 1 \): Maximum densification. The graph is basically a clique and the e grows quadratically with n.

\[
a = \frac{\gamma \log(d_{\text{max}}) + (2 - \gamma)\log(d_{\text{max}}) + o(1)}{\gamma \log(d_{\text{max}}) + (1 - \gamma)\log(d_{\text{max}}) + o(1)} = 2
\]
Case 2  **Evolving Degree Distribution**

• However, graphs with degree distribution $\gamma > 2$ can densify. Now assuming densification power law with exponent $a$, and the degree distribution can change over time
  
  – Q: How should the power-law degree exponent $\gamma$ change over time to allow for densification?
  
  – In fact, the degree distribution has to flatten over time to accumulate more mass in the tail as more nodes are added to allow for densification

• **Given a time evolving graph on $n$ nodes that evolves according to densification power law with exponent $a > 1$ and has power-law degree distribution $\gamma_n > 2$, then the degree exponent $\gamma_n$ evolves with the number of nodes $n$ as**

$$\gamma_n = \frac{4n^{a-1} - 1}{2n^{a-1} - 1}$$
Proof: Evolving Degree Distribution

• An undirected graph $G$ on $n$ nodes has edges $e = \frac{1}{2}nd\bar{d}$ where $\bar{d}$ is the average degree in $G$

• The densification power law exponent $a$ is

\[
a = \frac{\log(e)}{\log(n)} = \frac{\log(n) + \log(d\bar{d}) - \log(2)}{\log(n)}
\]

• In a graph with power-law degree distribution $p(x) = x^{-\gamma}$ with exponent $\gamma > 2$, the average degree $\bar{d}$ is

\[
\bar{d} \approx \int_{1}^{\infty} x p(x) \, dx = c \int_{1}^{\infty} x^{-\gamma + 1} \, dx = \frac{c}{2-\gamma} x^{-\gamma + 2} \bigg|_{1}^{\infty} = \frac{\gamma - 1}{\gamma - 2}
\]

• Substitute $\bar{d}$ in $a=\log(e)/\log(n)$ eq. and solve for $\gamma$

\[
\gamma_n = \frac{4n^{a-1} - 1}{2n^{a-1} - 1}
\]
Connections to Real-world Networks

Email Network

(a) Degree distribution

Constant Degree Exponent over Time

(b) Degree exponent over time

Physics Citation Network

(a) Degree distribution

Evolving Degree Distribution

(b) Degree exponent over time

1. Acquire some (more than one) social network datasets.

2. Finding some patterns from the dataset
   - Global patterns
     • About distribution or phenomenon (e.g. power-law about certain distribution, shrinking diameter)

3. Coming up with certain local models to explain the global patterns
   - Mathematical analysis
   - Simulation
Addition: Another Strategy for SNA Research

1. Setup a goal to achieve:
   – Finding or measuring something (e.g. community, centrality, abnormal nodes)
   – Summarization (e.g. network abstraction)
   – Prediction (e.g. node types or link types)

2. Finding some methods to do the above

3. Finding some datasets that more-or-less contain the gold standard to perform experiments.