A Basis for Systematic Analysis and Generation of Network Topologies

Priya Mahadevan

UC San Diego

Dmitri Krioukov (CAIDA), Kevin Fall (Intel Research),
Amin Vahdat (UC San Diego)
Importance of Network Topology

• Performance of protocols and applications
  • Routing, overlay networks
    For example: new routing protocol might offer $X$-time smaller routing tables for today but scale $Y$-time worse, with $Y >> X$

• Robustness of the network
• Traffic engineering
• Network management
• Spread of worms, etc.
Methodologies of Topology Research

- **Topology**
  - **Measurements, observations**
  - **Processes**
  - **Selection and abstraction**
    - Formalization
    - Network evolution modeling
    - Execution
  - **Comparison with the observed graphs against a set of important graph properties**
  - **Selection and abstraction**
    - Formalization
    - Network evolution modeling
    - Execution
  - **Synthetic ‘growing’ graphs**
  - **Synthetic ‘static’ graphs**
  - **Construction**
  - **Graph metrics to reproduce**
  - Extraction
  - **Observed graphs**
  - If graphs differ, refinements are needed: modify the set of reproduced graph metrics (on the left) or abstracted evolution rules (on the right)

- **Simulations**
Topology Evaluation Metrics

- Distance distribution
- Betweenness
- Clustering
- Assortativity coefficient / likelihood
- Spectrum

Problem in generating graphs?

- No known techniques to produce graphs with a given form of distance distribution, betweenness, etc.
- What if a new important metric is discovered?
Our Approach

Enumerable set of properties $P_d$, $d = 0, 1, \ldots$ that satisfy:

- **Constructibility**: construct graphs having these properties
- **Inclusion**: property $P_d$ subsumes $P_j$ where $j = 0, \ldots, d-1$
- **Convergence**: As $d$ increases, the set of graphs having property $P_d$ converges to the original graph $G$

Connectivity is the most basic property of network topologies

- We consider degree correlations among increasingly larger set of connected nodes
- $P_0$, $P_1$, $P_2$, $\ldots$, $P_n$ correspond to degree correlations among connected nodes of size $0$, $1$, $2$, $\ldots$, $n$
Outline

- Background
- Methodology
  - Graph generation
- Validation
  - AS-level graph (skitter)
  - Router-level graph (HOT)
- Limitations
- Conclusions
**Our $dK$-series**

$dK$-series: degree correlations within non-isomorphically simple connected subgraphs of size $d$

<table>
<thead>
<tr>
<th>Tag</th>
<th>Name</th>
<th>Subgraphs of size:</th>
<th>Symbolics</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0K$</td>
<td>Average node degree</td>
<td>0</td>
<td><img src="image" alt="Symbol1" /></td>
</tr>
<tr>
<td>$1K$</td>
<td>Node degree distribution</td>
<td>1</td>
<td><img src="image" alt="Symbol2" /></td>
</tr>
<tr>
<td>$2K$</td>
<td>Joint degree distribution</td>
<td>2</td>
<td><img src="image" alt="Symbol3" /></td>
</tr>
<tr>
<td>$3K$</td>
<td>Joint edge degree distribution</td>
<td>3</td>
<td><img src="image" alt="Symbol4" /></td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>$nK$</td>
<td>Full degree distribution</td>
<td>$n = # \text{ of nodes}$</td>
<td><img src="image" alt="Symbol5" /></td>
</tr>
</tbody>
</table>
More on $dK$-series

d$K$-series: degree correlations within non-isomorphic simple connected subgraphs of size $d$

How groups of $d$-nodes with different degrees interconnect

- $0K$ (average degree)
- $1K$ (degree distribution)
- $2K$

Connectivity between nodes having degrees $k_1$ and $k_2$

- $3K$

Connectivity among nodes having degrees $k_1$, $k_2$ and $k_3$
Example

- $0K$: avg deg = 2
- $1K$: $N(1)=1$, $N(2)=2$, $N(3)=1$
- $2K$: $N(1,3)=1$
  - $N(2,2)=1$
  - $N(2,3)=2$
- $3K$: $N_\Delta(2,2,3)=1$
  - $N_\Lambda(1,3,2)=2$

Compute probability distributions for each
**dK-graphs**

- Graphs associated with each $P_d$
  Original graph belongs to all subsets of $dK$-graphs ($d=0,..n$)
- As $d$ increases, $dK$ graphs converge to the original graph
- $dK$-random graph reproduces the specified property $P_d$, random with respect to all other properties
- What value of $d$ is sufficient for practical purposes to reproduce Internet topologies?
Constructing $dK$-random graphs

- Stochastic approach
  - In theory, can be generalized to any $d$; does not work well in practice

- Pseudograph (eg: PLRG)
  - In its original form, its only for 1K
  - Extended it for 2K, but not beyond

- Rewiring
  - Perturb the given graph (swap pairs of edges) such that property $P_d$ is preserved

  ![2K rewiring](image)

- Targeting Rewiring (Metropolis Dynamics)
  - Rewire graph such that it moves from reproducing $P_{d-1}$ to $P_d$
Outline

- Background
- Methodology
  - Graph generation
- Validation
  - AS-level graph (skitter)
  - Router-level graph (HOT)
- Limitations
- Conclusions
Validation

- We generated 0K, 1K, 2K and 3K-random graphs
- Compare our generated $dK$-random graphs with the original graph w.r.t. important topology metrics
- AS-level validation: skitter, BGP tables, WHOIS
- Router-level validation: HOT
Scalar metrics comparison for skitter

<table>
<thead>
<tr>
<th>Metric</th>
<th>0K</th>
<th>1K</th>
<th>2K</th>
<th>3K</th>
<th>skitter</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;k&gt;</td>
<td>6.31</td>
<td>6.34</td>
<td>6.29</td>
<td>6.29</td>
<td>6.29</td>
</tr>
<tr>
<td>r</td>
<td>0</td>
<td>-0.24</td>
<td>-0.24</td>
<td>-0.24</td>
<td>-0.24</td>
</tr>
<tr>
<td>&lt;C&gt;</td>
<td>0.001</td>
<td>0.25</td>
<td>0.29</td>
<td>0.46</td>
<td>0.46</td>
</tr>
<tr>
<td>d</td>
<td>5.17</td>
<td>3.11</td>
<td>3.08</td>
<td>3.09</td>
<td>3.12</td>
</tr>
<tr>
<td>σ_d</td>
<td>0.27</td>
<td>0.4</td>
<td>0.35</td>
<td>0.35</td>
<td>0.37</td>
</tr>
<tr>
<td>λ_1</td>
<td>0.2</td>
<td>0.03</td>
<td>0.15</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>λ_{n-1}</td>
<td>1.8</td>
<td>1.97</td>
<td>1.85</td>
<td>1.9</td>
<td>1.9</td>
</tr>
</tbody>
</table>
Distance Distribution in skitter

Distance in hops

PDF

0K–random
1K–random
2K–random
3K–random
Skitter
Clustering in skitter

![Graph showing clustering vs node degree](image)
$0K$-random graph

$1K$-random graph

$2K$-random graph

$3K$-random graph

Original graph (HOT)
## Scalar Metrics for HOT

<table>
<thead>
<tr>
<th>Metric</th>
<th>0K</th>
<th>1K</th>
<th>2K</th>
<th>3K</th>
<th>HOT</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;k&gt;</td>
<td>2.47</td>
<td>2.59</td>
<td>2.18</td>
<td>2.10</td>
<td>2.10</td>
</tr>
<tr>
<td>r</td>
<td>-0.05</td>
<td>-0.14</td>
<td>-0.23</td>
<td>-0.22</td>
<td>-0.22</td>
</tr>
<tr>
<td>&lt;C&gt;</td>
<td>0.002</td>
<td>0.009</td>
<td>0.001</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>d</td>
<td>8.48</td>
<td>4.41</td>
<td>6.32</td>
<td>6.55</td>
<td>6.81</td>
</tr>
<tr>
<td>σ&lt;sub&gt;d&lt;/sub&gt;</td>
<td>1.23</td>
<td>0.72</td>
<td>0.71</td>
<td>0.84</td>
<td>0.57</td>
</tr>
<tr>
<td>λ&lt;sub&gt;1&lt;/sub&gt;</td>
<td>0.01</td>
<td>0.034</td>
<td>0.005</td>
<td>0.004</td>
<td>0.004</td>
</tr>
<tr>
<td>λ&lt;sub&gt;n-1&lt;/sub&gt;</td>
<td>1.989</td>
<td>1.967</td>
<td>1.996</td>
<td>1.997</td>
<td>1.997</td>
</tr>
</tbody>
</table>
Distance Distribution in HOT

![Graph showing distance distribution in HOT]

- **PDF**: Probability Density Function
- **Distance in hops**: X-axis
- **0K-random**, **1K-random**, **2K-random**, **3K-random**, **HOT**: Different data sets plotted
Betweenness in HOT

Normalized node betweenness vs. Node degree

- HOT
- 0K-random
- 1K-random
- 2K-random
- 3K-random
Results summary

- For AS-level topologies:
  - $dK$-series convergence is fast
  - $2K$-random graphs capture most metrics; need $3K$ for clustering

- For router-level topologies:
  - $dK$-series convergence is slower
  - $3K$-random graphs reproduce most metrics

- Router-level topologies are less random; reflect careful design and planning
**Discussions and Limitations**

- For graphs that we considered, $d=3$ seems sufficient.
- Not all graphs can be approximated using $3K$.
- What if we have to reproduce a new metric?
  - Increase $d$.
  - Extreme case of $d = n$, generated graphs must be isomorphic to the given graph.
  - Computational complexity grows rapidly with $d$.
- Cannot discover evolutionary growth of a network.
Conclusions

- No need to capture individual metrics
- Our $dK$-series, $d = 0, 1, ...n$ specify degree correlations within non-isomorphic simple connected subgraphs of size $d$
- By increasing $d$, we capture more complex properties of a given graph
- $d=3$ is sufficient to reproduce important metrics of observed Internet graphs
More Information

1. A Basis for Systematic Analysis of Network Topologies,
   *Priya Mahadevan*, *Dmitri Krioukov*, *Kevin Fall*, and *Amin Vahdat*. *ACM SIGCOMM 2006.*