The Role of Graph Entropy and Constraints on Fault Localization and Structure Evolution in Data Networks

Mathematics of Networks 15, University of Bath

Phil Tee
e-mail: phil@moogsoft.com
September 23, 2016

University of Sussex and Moogsoft Inc
The Journey From Fault Localization to Graph Evolution

Graph Entropy and Fault Localization

Vertex Entropy and Significant Events [1]

Deviation of Degree Distributions from Power Laws

Entropic Model of Graph Evolution

Constraint Models of Graph Evolution
• Fault localization challenge = too many noisy events
• Graph entropy could eliminate noise, but is global
• Introduce local vertex entropy following Dehmer [2].
• Measures applied to real datasets
• Problems with power law node degree fits with these networks
• Introduce a new constraint model to explain deviations
• New vertex entropy, network growth model possible?
Problem Statement
Spotting Events that Threaten Availability

- Fault Localization Algorithms most common solution
- They struggle to scale to 10,000’s events per second, mostly noise
- Common Approaches to Mitigate
  - Manual blacklisting
  - Restriction of monitoring to core devices
  - Deploy headcount
- Can we use topology and graph theory here? ¹

“74% Application Incidents reported by End Users”

¹We use standard Graph Theory notation throughout, see any standard text such as [3]. We denote a graph by $G(V, E)$ a double set of vertices $V$ and their edges $E$
Theoretical Background
How important is a Node in a Graph

- Network Science demonstrates some nodes are more critical (Barabási-Albert [4], [5])
- High degree nodes destroy graph connectivity quicker than low degree
- Conclusion: **High Degree Nodes are more "Important"**
- Many other measures exist, we focus on entropy
What is Graph Entropy $H[G(V, E)]$

- Measures structural *information* in a graph. The more meshed a graph, the lower the entropy
- **Chromatic Entropy**
  - Defined using Chromatic number of the graph. Acts like "negentropy"
- **Körner or Structural Entropy**
  - Closely related, uses non adjacent sets of vertices
- **Von Neumann Entropy**
  - Defined by the eigenvalues of the *Laplacian* matrix of a graph. Measures the connectivity of a graph

---

2We assume in our treatment that vertex emission probabilities are all uniform
Graph Entropy a Measure of Redundancy

Graph Types that Maximise and Minimise Entropy

<table>
<thead>
<tr>
<th></th>
<th>Chromatic</th>
<th>Structural</th>
<th>Von Neumann</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum</td>
<td>$K_n$</td>
<td>$S_n$</td>
<td>$K_n$</td>
</tr>
<tr>
<td>Minimum</td>
<td>$S_n$</td>
<td>$K_n$</td>
<td>$P_n$</td>
</tr>
</tbody>
</table>

\(^3\text{In all of our work we only consider connected, simple graphs}\)
Global Measures Computationally too Complex

- It is only valid globally, no value for an individual node
- All are expensive to compute, and contain \( NP \) complete problems

We need a vertex value such that

\[
H[G(\mathcal{V}, \mathcal{E})] \sim \sum_{v \in \mathcal{G}} H(v)
\]
• Dehmer ([2]) creates a framework for calculating graph entropy in terms of vertices

• Introduces vertex information functional $f_i(v)$ of a node $v$, with vertex **probability** defined as

$$p_i(v) = \frac{f_i(v)}{\sum_{v \in G} f_i(v)}$$

• Node entropy $H(v) = -p_i \log p_i$, and total graph entropy $H(G) = \sum_{v \in G} H(v)$
Introducing the Local Vertex Entropy $\text{VE}$ and $\text{VE'}$

- We define an inverse degree entropy for a node $\text{VE}(v)$ as:
  \[
p_i(v_i) = \frac{1}{k_i} \text{ where } k_i \text{ is the degree of } v_i, \quad \text{VE}(v_i) = \frac{1}{k_i} \log_2(k_i)
  \]

- And fractional degree entropy of a node $\text{VE'}(v)$ as:
  \[
p_i(v_i) = \frac{k_i}{2|E|}, \quad \text{VE'}(v_i) = \frac{k_i}{2|E|} \log_2 \left( \frac{2|E|}{k_i} \right)
  \]

- These two measures do not take into account high degree nodes which are redundantly connected into the graph
Not all High Degree Nodes are Equal!

- To capture importance more accurately we suppress entropy for highly meshed nodes.
- A highly meshed network has local similarity to the perfect graph $K_n$. The modified clustering coefficient $C_i$ of the neighborhood of a vertex $i$ scales our metrics as:

$$C_i = \frac{2|E_1(v_i)|}{k_i(k_i + 1)}, \quad NVE(v) = \frac{1}{C_i} VE(v) \text{ and } NVE'(v) = \frac{1}{C_i} VE'(v)$$

- And for the whole graphs:

$$NVE(G) = \sum_{i=0}^{i<n} \frac{(k_i + 1)}{2|E_i|} \log_2(k_i)$$

$$NVE'(G) = \sum_{i=0}^{i<n} \frac{k_i^2(k_i + 1)}{4|E||E_i|} \log_2 \left( \frac{2|E|}{k_i} \right)$$

---

$^4$we include the central vertex in our version to avoid problematic zeros.
Comparing \(NVE\) and \(NVE'\) to Global Entropy Measures

Values of Normalized Entropy for Special Graphs

<table>
<thead>
<tr>
<th>(S_n)</th>
<th>(NVE)</th>
<th>(NVE')</th>
</tr>
</thead>
<tbody>
<tr>
<td>(S_n)</td>
<td>(\frac{n}{2(n-1)} \log_2(n - 1))</td>
<td>(\frac{1}{2} \log_2{2(n - 1)} + \frac{n}{4})</td>
</tr>
<tr>
<td>(K_n)</td>
<td>(\frac{n}{n-1} \log_2(n - 1))</td>
<td>(\log_2(n))</td>
</tr>
<tr>
<td>(P_n)</td>
<td>(\frac{3}{4}(n - 2))</td>
<td>(\frac{1}{n-1} + \frac{3n-4}{2(n-1)} \log_2(n - 1))</td>
</tr>
<tr>
<td>(C_n)</td>
<td>(\frac{3}{4}n)</td>
<td>(\frac{3}{2} \log_2(n))</td>
</tr>
</tbody>
</table>

Maximal and Minimal Total Vertex Entropy Graph Types

<table>
<thead>
<tr>
<th>(NVE)</th>
<th>(NVE')</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum</td>
<td>(C_n \sim P_n)</td>
</tr>
<tr>
<td>Minimum</td>
<td>(S_n)</td>
</tr>
</tbody>
</table>

Close inspection of the minima and maxima indicate that \(NVE'\) has similar limit behavior to Structural entropy, and \(NVE\) to Chromatic entropy.
Our Experimental Data Sets

- **Commercial:** Moogsoft routinely collects the following from our customers
  - **Topology:** Manually curated and automatically discovered lists of node to node connections
  - **Network Events:** From their (Moogsoft Supplied) management systems collections of monitored network events
  - **Incidents:** From their help-desk and escalation systems collections of escalated events.

- Our principal dataset covers **225,239 nodes, 96,325,275 events and 37,099 incidents**

- **Academic:** "The Internet Topology Zoo" by S.Knight *et al* [6] curates **15,523 network nodes** across a number of real world telecoms networks.
What Would an Ideal Distribution Look Like?

Ideal Distribution of Incidents and Events
Distributions of $NVE$

Distribution of Incidents and Events by $NVE$

Analysis

Noticeable separation of distribution to favor high $NVE$ for Incidents
Distributions of $NVE'$

**Analysis**

Separation of distribution of Incidents and Events by $NVE'$ statistically significant
Comparing $NVE'$ to Degree Importance

- Taking the same dataset and comparing distributions it is evident $NVE'$ is more predictive than node degree

Distribution of Incidents by $NVE'(v)$ and Node Degree

Distribution of Events by $NVE'(v)$ and Node Degree
A Constraint Based model of Network Growth
The Barabási-Albert Model Recap

- Deduces a degree distribution power law from the principles of
  - *Growth*: Starting at time $t_i$ a single new node is added at each time interval $t$ to a network of $m_0$ nodes. When the node is added to the network it attaches to $m$ other nodes. This process continues indefinitely.
  - *Preferential Attachment*: The node attaches to other nodes with a probability determined by the degree of the target node, such that more highly connected nodes are preferred over lower degree nodes.
- Central prediction is power law degree distributions:
  \[ P(k) = \frac{2m^2 t}{m_0 + t} \frac{1}{k^\gamma}, \text{ with } \gamma = 3 \]
Analysis of Networks Demonstrates Deviations at High $k$

- Considerable deviations from Power Law distribution at high degree in Analyzed Networks [7], [6]

IT Zoo Networks

Facebook Friendship Graph
Not Confined to any Particular Type of Graph...

- Seen again in citation networks, and across all 20 datasets analyzed.

Patent Citations

\[ \gamma : 2.615-3.118 \]

\[ <\gamma> : 2.981 \]

ArXiv HepTh Citation Graph

\[ \gamma : 2.709-2.142 \]

\[ <\gamma> : 2.220 \]
Our Proposed Model

- We propose a simple average constraint on the maximum degree of a node ‘c’
- We scale the probability of attachment by the capacity of the node relative to average capacity

\[ \Pi_i = \zeta_i \times P(attachment) \text{, where } \zeta_i = \frac{(c - k_i)}{\langle c_i(t) \rangle} \]

- This can be solved using the continuum approach for \( P(k) \)

\[
P(k) = \frac{2c\rho^{2/\alpha} t}{\alpha(t + m_0)} \left( \frac{(c - k)^{\frac{2}{\alpha} - 1}}{k^{\frac{2}{\alpha} + 1}} \right) \sim \frac{1}{k\gamma}
\]

where \( \alpha = \frac{c}{c - 2m} \) and \( \rho = \frac{m}{c - m} \) and \( \gamma = \frac{2}{\alpha} + 1 \)
Analysis of Datasets Confirms an Improved Prediction for $\gamma$

- 9 of 20 datasets analyzed show an agreement with calculated $\gamma$ of < 10% (shown in bold below)

<table>
<thead>
<tr>
<th>Source</th>
<th>$\gamma$ Calculated</th>
<th>$\gamma$ Measured</th>
<th>$\delta$ Constraints</th>
<th>$\delta$ Preferential</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arxiv - HepTh (Cit)</td>
<td>2.71</td>
<td>2.71</td>
<td>0.05%</td>
<td>11.00%</td>
</tr>
<tr>
<td>Berkley Stanford Web</td>
<td>2.89</td>
<td>2.85</td>
<td>1.44%</td>
<td>5.00%</td>
</tr>
<tr>
<td>IT Zoo</td>
<td>2.50</td>
<td>2.54</td>
<td>1.62%</td>
<td>18.00%</td>
</tr>
<tr>
<td>Pokec</td>
<td>2.70</td>
<td>2.65</td>
<td>1.68%</td>
<td>13.00%</td>
</tr>
<tr>
<td>Web Provider</td>
<td>2.78</td>
<td>2.68</td>
<td>3.54%</td>
<td>12.00%</td>
</tr>
<tr>
<td>IMDB Movie Actors</td>
<td>2.43</td>
<td>2.30</td>
<td>5.83%</td>
<td>30.43%</td>
</tr>
<tr>
<td>Arxiv - HepTh (Collab)</td>
<td>2.81</td>
<td>2.64</td>
<td>6.42%</td>
<td>13.00%</td>
</tr>
<tr>
<td>Internet Router</td>
<td>2.66</td>
<td>2.48</td>
<td>7.15%</td>
<td>20.97%</td>
</tr>
<tr>
<td>Arxiv - Astro Phys</td>
<td>2.54</td>
<td>2.77</td>
<td>8.31%</td>
<td>8.00%</td>
</tr>
<tr>
<td>Twitter (Follower)</td>
<td>2.96</td>
<td>2.65</td>
<td>11.54%</td>
<td>13.00%</td>
</tr>
<tr>
<td>Patent Citation</td>
<td>2.59</td>
<td>2.28</td>
<td>13.9%</td>
<td>32.00%</td>
</tr>
<tr>
<td>Co-authors, math</td>
<td>2.87</td>
<td>2.5</td>
<td>14.80%</td>
<td>20.00%</td>
</tr>
<tr>
<td>Enron Email</td>
<td>2.96</td>
<td>2.57</td>
<td>15.05%</td>
<td>17.00%</td>
</tr>
<tr>
<td>AS Skitter</td>
<td>2.92</td>
<td>2.47</td>
<td>18.19%</td>
<td>21.00%</td>
</tr>
<tr>
<td>Arxiv - Cond Matt</td>
<td>2.69</td>
<td>2.24</td>
<td>20.37%</td>
<td>34.00%</td>
</tr>
<tr>
<td>Metabolic, <em>E. coli</em></td>
<td>2.73</td>
<td>2.20</td>
<td>24.13%</td>
<td>36.36%</td>
</tr>
<tr>
<td>Twitter (Circles)</td>
<td>2.72</td>
<td>2.01</td>
<td>35.44%</td>
<td>49.00%</td>
</tr>
<tr>
<td>Co-authors, neuro</td>
<td>2.88</td>
<td>2.1</td>
<td>37.36%</td>
<td>42.86%</td>
</tr>
<tr>
<td>Facebook</td>
<td>2.25</td>
<td>1.39</td>
<td>62.08%</td>
<td>116.00%</td>
</tr>
<tr>
<td>Co-authors, SPIRES</td>
<td>2.37</td>
<td>1.2</td>
<td>97.58%</td>
<td>150.00%</td>
</tr>
</tbody>
</table>
Could Graph Entropy Explain Both Growth Models?
Basis of Model

• The 2\textsuperscript{nd} law of thermodynamics states that total entropy must tend to a maximum in any closed system.
• One consequence is the concept of entropic force, which explains natural processes such as osmosis.

\[ F = T \Delta S \]

\( T \) is thermodynamic ‘temperature’ and \( S \) is the entropy of the system.

• Entropy of the whole graph has been considered before ([8]). We imagine a vertex level dynamic process.
• Propose probability of attachment to a given node is proportional to the relative ‘attraction’ of ‘force’ exerted by a particular node:

\[ \Pi_i = \frac{F_i(v_i)}{\sum_{j \neq i} F_j(v_j)} \]

we can then follow continuum analysis.
Overview of Analysis

- We factor out the temperature dependence, and by approximating the denominator using an expectation value of the change in entropy, we arrive at

\[ \Pi_i = \epsilon \Delta S_i, \text{ where } \epsilon = \frac{1}{|V| \times \mathbb{E}(\Delta S)} \]

- To calculate \( \Delta S_i \) we note \( \Delta S_i = \frac{\partial S_i}{\partial k} \times \delta k \), with, for a single time step, \( \delta k = 2m \). This gives as an attachment probability

\[ \Pi_i = \epsilon 2m \frac{\partial S_i}{\partial k} \]

- For \( S_i \) we can insert our previous definition of \( NVE' \), approximating the clustering coefficient to yield

\[ S_k = \frac{k^2}{4m^2t^2} \log \left( \frac{2mt}{k} \right) \]
We derive the final form of the degree evolution partial differential equation to be

\[ \frac{\partial k}{\partial t} = 2m \Pi_i = -\epsilon \frac{k}{t} \left\{ \frac{1}{2} + \log \left( \frac{k}{2mt} \right) \right\} \]

Which as \( k \ll 2mt \) we can expand the logarithm to obtain

\[ \frac{\partial k}{\partial t} \approx \frac{\epsilon k}{2t} - \frac{\epsilon k^2}{2mt} + \epsilon O \left( \frac{k}{2mt} \right)^2 \]
• Taylor series expansion has preferential attachment and constraints as the first two terms!
• Higher terms may reveal even more complex corrections to scale freedom
• The constant $\epsilon$ explains why $\gamma$ is never exactly 3 even at low $k$
• The model has been arrived at from fundamental principles and could explain why nodes preferentially attach
Conclusions
Conclusions

- Vertex entropy, $NVE'$ is useful at eliminating noisy events
- The constraints model more accurately matches real network metrics
- Vertex entropy can be used to build an entropic model of network growth.
- This model has constraints emerging naturally and explains why $\gamma$ is never exactly 3
I would like to thank my colleagues at Moogsoft and our customers who gave us their time and data. I would also like to thank Ian and George for the continual help, suggestions, and support in developing these ideas. Finally, I would like to thank Christine for her debate, encouragement, and editorial services given freely!


