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

Mathematics of Networks 15, University of Bath

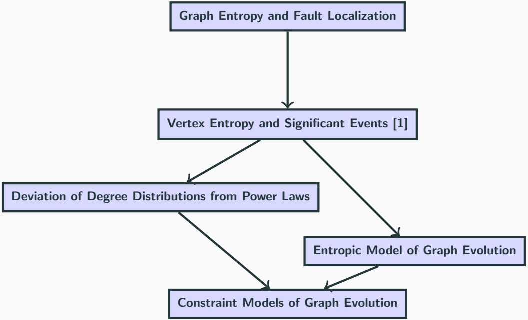
Phil Tee
email: phil@moogsoft.com

September 23, 2016

University of Sussex and Moogsoft Inc

Overview of our Research

The Journey From Fault Localization to Graph Evolution



Summary

- 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

The Problem of Event Overload in Event Management

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"

FORRESTER

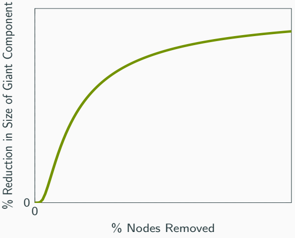
[1] We use standard Graph Theory notation throughout, see any standard text such as [3]. We denote a graph by $G(V, E)$ a double set of vertices V and their edges E

Theoretical Background

Graph Structure and Node Importance

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



Graph Entropy and Network Resilience

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**[2]
  - 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

---

[2]We assume in our treatment that vertex emission probabilities are all uniform

Graph Entropy a Measure of Redundancy

S_4 K_4 C_4 P_4

Graph Types that Maximise and Minimize Entropy[3]

	Chromatic	Structural	Von Neumann
Maximum	K_n	S_n	K_n
Minimum	S_n	K_n	P_n

[3] 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 $H[G(V,E)] \sim \sum_{v \in G} H(v)$

The Dehmer Vertex Informational Functional

- 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

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

- 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 VE and VE'

- We define an inverse degree entropy for a node $VE(v)$ as:

$$\boxed{p_i(v_i) = \frac{1}{k_i} \text{ where } k_i \text{ is the degree of } v_i,\ VE(v_i) = \frac{1}{k_i}\log_2(k_i)}$$

- And fractional degree entropy of a node $VE'(v)$ as:

$$\boxed{p_i(v_i) = \frac{k_i}{2|E|},\ VE'(v_i) = \frac{k_i}{2|E|}\log_2\left(\frac{2|E|}{k_i}\right)}$$

- 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 [4] clustering coefficient C_i of the neighborhood of a vertex i scales our metrics as:

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

- And for the whole graphs:

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

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

[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

	NVE	NVE'
S_n	$\frac{n}{2(n-1)}\log_2(n-1)$	$\frac{1}{2}\log_2\{2(n-1)\}+\frac{n}{4}$
K_n	$\frac{n}{n-1}\log_2(n-1)$	$\log_2(n)$
P_n	$\frac{3}{4}(n-2)$	$\frac{1}{n-1}+\frac{3n-4}{2(n-1)}\log_2(n-1)$
C_n	$\frac{3}{4}n$	$\frac{3}{2}\log_2(n)$

Maximal and Minimal Total Vertex Entropy Graph Types

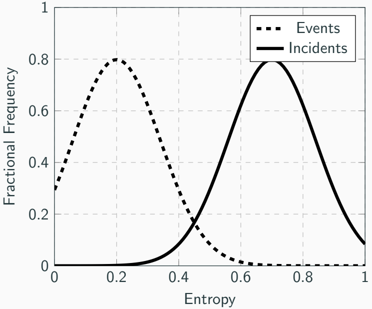
	NVE	NVE'
Maximum	$C_n \sim P_n$	S_n
Minimum	S_n	K_n

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

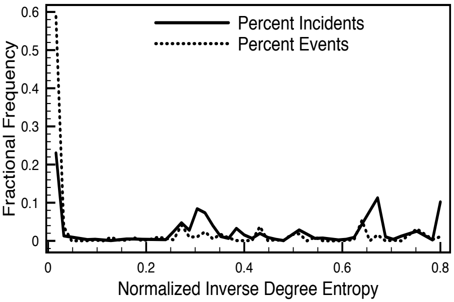
- **Commercial:** Moogsoft routinely collects the following from our customers
  - **Topology:** Manually curated and automatically discovered lists of node to node connections
  - **Network Events:** From their (Moogsoft Supplied) management systems collections of monitored network events
  - **Incidents:** 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 and Ideal Distribution Look Like?



Ideal Distribution of Incidents and Events

Distributions of *NVE*

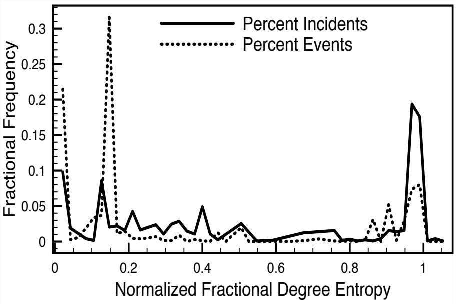


Distribution of Incidents and Events by *NVE*

Analysis

Noticeable separation of distribution to favor high *NVE* for Incidents

Distributions of NVE'



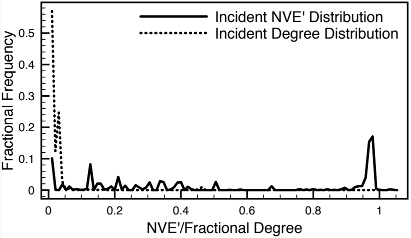
Distribution of Incidents and Events by NVE'

Analysis

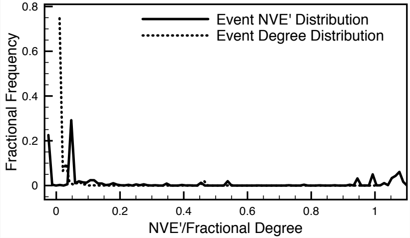
Separation of distribution of Incidents and Events by NVE' statistically significant

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

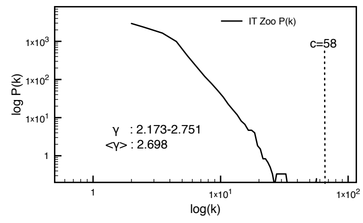
The Barabási-Albert Model Recap

- 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:

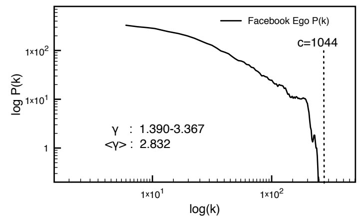
$$\boxed{P(k) = \frac{2m^2 t}{m_0 + t}\frac{1}{k^\gamma}, \text{ with } \gamma = 3}$$

Analysis of Networks Demonstrates Deviations at High k

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



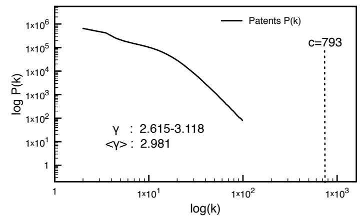
IT Zoo Networks



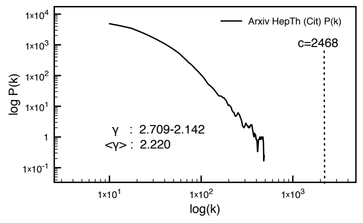
Facebook Friendship Graph

Not Confined to any Particular Type of Graph...

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



Patent Citations



ArXiv HepTh Citation Graph

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 $P(k)$

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

$$\text{where } \boxed{\alpha = \frac{c}{c - 2m} \text{ and } \rho = \frac{m}{c - m} \text{ and } \gamma = \frac{2}{\alpha} + 1}$$

Analysis of Datasets Confirms an Improved Prediction for γ

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

Comparison of γ Predictions Between Preferential Attachment and Constraints Model

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

Could Graph Entropy Explain Both Growth Models?

Basis of Model

- The 2^{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.

$$F = \mathbf{T}\Delta\mathbf{S}$$

  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:

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

  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

$$\Pi_i = \epsilon \Delta \mathbf{S}_i, \text{ where } \epsilon = \frac{1}{|V| \times \mathbb{E}(\Delta \mathbf{S})}$$

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

$$\boxed{\Pi_i = \epsilon 2m \frac{\partial \mathbf{S}_i}{\partial k}}$$

- For $\mathbf{S}_i$ we can insert our previous definition of NVE', approximating the clustering coefficient to yield

$$\boxed{S_k = \frac{k^2}{4m^2 t^2} \log\left(\frac{2mt}{k}\right)}$$

Approximating S_i and Final Form for Degree Evolution

- We derive the final form of the degree evolution partial differential equation to be

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

- Which as $k \ll 2mt$ we can expand the logarithm to obtain

$$\boxed{\frac{\partial k}{\partial t} \approx \frac{\epsilon k}{2t} - \frac{\epsilon k^2}{2mt} + \epsilon O\left(\frac{k}{2mt}\right)^2}$$

Discussion

- 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 ϵ explains why γ 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 γ is never exactly 3

Acknowledgements

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!

Summary References

References I

[1] P. Tee, G. Parisis, and I. Wakeman, "Towards an Approximate Graph Entropy Measure for Identifying Incidents in Network Event Data," no. AnNet, pp. 1049–1054, 2016.

[2] M. Dehmer, "Information processing in complex networks: Graph entropy and information functionals," *Applied Mathematics and Computation*, vol. 201, no. 1-2, pp. 82–94, 2008.

[3] B. Bollobàs, *Modern Graph Theory*. Springer-Verlag New York, 1998.

[4] R. Albert and A.-L. Barabási, "Statistical mechanics of complex networks," *Review of Modern Physics*, vol. 74, no. January, 2002.

[5] R. Albert, H. Jeong, and A. Barabasi, "Error and attack tolerance of complex networks," *Nature*, vol. 406, no. 6794, pp. 378–82, 2000.

References II

[6] S. Knight, H. X. Nguyen, N. Falkner, R. Bowden, and M. Roughan, "The internet topology zoo," *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 9, pp. 1765–1775, 2011.

[7] J. Leskovec and A. Krevl, "SNAP Datasets: Stanford Large Network Dataset Collection," url{http://snap.stanford.edu/data}, jun 2014.

[8] J. Park and M. E. J. Newman, "Statistical mechanics of networks," *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, vol. 70, no. 6 2, pp. 1–13, 2004.