The Role of Graph Entropy in Fault Localization and Network Evolution

Philip Tee

Supervisors: Prof. Ian Wakeman.
Dr George Parisis

Department of Informatics
University of Sussex

This dissertation is submitted for the degree of
Doctor of Philosophy

November 2017

I dedicate this thesis to the memory of my father Arthur Tee who passed away during my studies. Dad, all of those hours spent talking about science gave me the bug. It turns out the left hand rule isn't quite all there is to know...

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this thesis are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This thesis is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements.

This thesis conforms to a 'papers style' format in which a substantial part of the contents are reproductions from published and un-published submissions to journals. In particular chapters 4, 5 and 6 contain publications submitted to peer reviewed journals of which I was the first author. I hereby declare that substantially all of the expermiental and theoretical work presented in those papers was my own, except as detailed in the preamble to the papers in those chapters.

Philip Tee
November 2017

Acknowledgements

I would like to thank all of the people that have offered help and encouragement during my studies. Firstly my wife Christine, who first planted the seed I could do this and finally scratch the itch! I am certain that without her challenging conversations and gentle prodding I would never have made it through the process...

I would also like to thank Prof. David Weir who first suggested I work with Prof. Ian Wakeman, who has been a tremendous mentor.

My two advisors, Ian and Dr George Parisis have been constant companions on the journey and have patiently read through early drafts and ideas. Thank you for passing on the 'craft' of research.

I have had numerous conversations, suggestions and constructive criticism from Dr István Z. Kiss, Dr Luc Berthouze, Dr Justin Coon, Dr Richard Clegg, and Prof. Jonathan Dawes. These have helped me shape my thoughts into the work presented in this thesis.

I would also like to mention my lifelong friend Dr Michael Bluck with whom I first shared the love of maths and science. He has watched with some amusement whilst I made my way through the doctoral studies, a path he started at the same time as I, but finished 25 years earlier!

Finally I would like to thank the examiners for invaluable comments and suggestions regarding the proofs, which have contributed to an improved manuscript.

Abstract

The design of a communication network has a critical impact on its effectiveness at delivering service to the users of a large scale compute infrastructure. In particular, the reliability of such networks is increasingly vital in the modern world, as more and more of our commercial and social activity is conducted using digital platforms. Systems to assure service availability have been available since the emergence of Mainframes, with the System 360 in 1964, and although commercially widespread, the scientific understanding is not as deep as the problem warrants. The basic operating principle of most service assurance systems combines the gathering of status messages, which we term as events, with algorithms to deduce from the events where potential failures may be occurring. The algorithms to identify which events are causal, known as root cause analysis or fault localization, usually rely upon a detailed understanding of the network structure in order to determine those events that are most helpful in diagnosing and remediating a service threatening problem. The complex nature of root cause algorithms introduces scalability limits in terms of the number of events that can be processed per second. Unfortunately as networks grow, the volume of events produced continues to increase, often dramatically.

The dependence of root cause analysis algorithms on network structure presents a significant challenge as networks continue to grow in scale and complexity. As a consequence of this, and the growing reliance upon networks as part of the key fabric of the modern economy, the commercial importance and the scale of the engineering challenges are increasing significantly.

In this thesis I outline a novel approach to improving the scalability of event processing using a mathematical property of networks, graph entropy. In the first two papers described in this thesis, I apply an efficiently computable approximation of graph entropy to the problem of identifying important nodes in a network. In this context, importance is a measure of whether the failure of a node is more likely to result in a significant impact on the overall connectivity of the network, and therefore likely to lead to an interruption of service. I show that by ignoring events from unimportant network nodes it is possible to significantly reduce the event rate that a root cause algorithm needs to process. Further, I demonstrate that

unimportant nodes produce very many events, but very few root causes. The consequence is that although some events relating to root causes are missed, this is compensated for by the reduction in overall event rate. This leads to a significant reduction of the event processing load on management systems, and therefore increases the effectiveness of current approaches to root cause analysis on large networks.

Analysis of the topology data used in the first two papers revealed interesting anomalies in the degree distribution of the network nodes. This motivated the later focus of my research to investigate how graph entropy and network design considerations could be applied to the dynamical evolution of networks structures, most commonly described using the *Preferential Attachment* model of Barabási and Albert. A common feature of a communication network is the presence of a constraint on the number of logical or physical connections a device can support. In the last of the three papers in the thesis I develop and present a constrained model of network evolution, which demonstrates better quantitative agreement with real world networks than the preferential attachment model. This model, developed using the continuum approach, still does not address a fundamental question of random networks as a model of network evolution. Why should a node's degree influence the likelihood of it acquiring connections? In the same paper I attempt to answer that question by outlining a model that links vertex entropy to a node's attachment probability. The model successfully reproduces some of the characteristics of preferential attachment, and illustrates the potential for entropic arguments in network science.

Put together, the two main bodies of work constitute a practical advance on the state of the art of fault localization, and a theoretical insight into the inner workings of dynamic networks. They open up a number of interesting avenues for further investigation.

Table of contents

List of figures **xiii**

List of tables **xv**

Nomenclature **xvii**

1 Introduction **1**

1.1 Overview 1

1.1.1 Summary 1

1.2 Fault Localization Background 2

1.2.1 Overview of Fault Localization 2

1.2.2 Algorithmic Approaches 4

1.2.3 Known Limitations of Fault Localization 12

1.2.4 Summary of Research Activities 13

2 Graph Theory Themes **17**

2.1 Graph Theory Essentials 17

2.1.1 Basic Definitions 17

2.1.2 Adjacency, Coloring and Clustering 19

2.2 Graph Entropy 21

2.2.1 Körner or Structural Entropy 22

2.2.2 Chromatic Entropy 24

2.2.3 Alternative Formulations 25

2.3 Dynamic Graphs and Graph Evolution 25

2.3.1 Random Graphs 25

2.3.2 The Preferential Attachment Model 26

2.3.3 Extensions to Preferential Attachment 27

3 Overview of Published Work **31**
3.1 Towards and Approximate Graph Entropy Measure for Identifying Incidents in Network Event Data . . . 31
3.2 Vertex Entropy as a Critical Node Measure in Network Monitoring . . . 31
3.3 Constraints and Entropy in a Model of Network Evolution . . . 32
3.4 Other Submissions and Conference Talks . . . 32

4 Towards and Approximate Graph Entropy Measure for Identifying Incidents in Network Event Data **35**
4.1 Background to First Publication . . . 35
4.1.1 Motivation and Summary of Contribution . . . 35
4.1.2 Theoretical Contribution . . . 35
4.1.3 Data and Methods Used . . . 36
4.1.4 Contributions from Co-Authors . . . 37
4.1.5 Related Work . . . 37
4.2 Discussion . . . 45
4.2.1 The General Application of Vertex Entropy as a Root Cause Indicator 45
4.2.2 Post Publication Perspective . . . 45

5 Vertex Entropy as a Critical Node Measure in Network Monitoring **47**
5.1 Background to Second Publication . . . 47
5.1.1 Motivation and Summary of Contribution . . . 47
5.1.2 Theoretical Contribution . . . 47
5.1.3 Data and Methods Used . . . 48
5.1.4 Contributions from Co-Authors . . . 48
5.1.5 Related Work . . . 49
5.2 Discussion . . . 67
5.2.1 Comparing Vertex Entropy and Graph Entropy of Sampled Graphs . 67
5.3 F_β Analysis and ROC curves . . . 69
5.3.1 F_β Analysis with β=100 . . . 69
5.3.2 ROC Curve Analysis . . . 70
5.3.3 Corrected Proof of Proposition 1 . . . 72
5.3.4 Normalization of Inverse Degree Entropy . . . 74
5.3.5 Shannon's Desiderata . . . 75
5.3.6 Definition of the j-Sphere . . . 75

6 Constraints and Entropy in a Model of Network Evolution 81
6.1 Background to Third Publication 81
6.1.1 Motivation and Summary of Contribution 81
6.1.2 Theoretical Contribution 82
6.1.3 Data and Methods Used 82
6.1.4 Contributions from Co-Authors 83
6.1.5 Related Work 83
6.2 Discussion 102

7 Conclusion and Future Directions 105

References 107

Appendix A Personal Biography 115

List of figures

1.1 Steinder - Sethi Ontology of Fault Localization Techniques 4
1.2 Example of a Basic Rules Based System 7
1.3 A Simple Codebook Causality Graph 9
1.4 A Simple Downstream Topology 10
1.5 The Journey From Fault Localization to Graph Evolution 15

2.1 An Example Graph 18
2.2 Special Graphs of Order Four 19

5.6 A Counter Example Graph to the Claim in Proposition 1. 72
5.1 Sampled sum of Vertex Entropies for $G(N,p)$ Erdős-Rényi Graphs, with $p \in [0.3, 0.7]$ and $|V| = 100$ 76
5.2 Sampled sum of Vertex Entropies for Scale Free Graphs with $m \in [2,23]$ and $|V| = 300$ 77
5.3 Sampled sum of Vertex Entropies for Scale Free Graphs with for $|V| \in [70,270]$, and edge density 94-98% 78
5.4 NVE'(v) F_β Plots against Recall 79
5.5 ROC Curve Analysis for Event and Incident Distribution for each Vertex Entropy Metric 80

List of tables

1.1 Description of Type and Content of Typical Event Attributes 7
1.2 Codebook for Causality Graph in Figure 1.3 8
1.3 Codebook for Figure 1.4 10

5.1 Calculation of ROC Components in Event and Incident Data 71
5.2 Area Under the ROC Curve (AUC) Analysis of Entropy Measures 72

Nomenclature

Roman Symbols

$\mathbf{A}_{ij}$ The Adjacency matrix of a Graph G

c Maximum degree of a graph node

C_n The Cycle Graph over n Vertices

C_i^j The Clustering coefficient of node i in a j-Sphere

$\mathbf{D}_{ij}$ The Degree matrix of a Graph G

$d_H(\mathbf{A},\mathbf{B})$ Hamming distance between two vectors $\mathbf{A}$ and $\mathbf{B}$

E Set of the edges of a graph

F_β Modified F_β score comparing precision and recall of a classification algorithm

$G(V,E)$ A graph, consisting of the set of vertices V and the set of edges between them E

$H(G,P)$ The Structural or Körner Entropy of a Graph G and Probability Distribution $P(v)$ over the set of vertices $v \in V$

$I_c(G)$ The Chromatic Entropy of Graph G

k Degree of a graph node

K_n The Complete or Perfect Graph of n Vertices

$\mathbf{L}_{ij}$ The Laplacian matrix of a Graph G

m Number of nodes a new node connects to in dynamic network models

m_0 Initial size of the network in dynamic network models

P_n The Path Graph over n Vertices

$P(k)$ The Degree Probability Distribution Function

S_n The Star Graph of n Vertices

S_i^j The j-Sphere of node i, i.e. all nodes within j hops of node i

V Set of the vertices of a graph

Greek Symbols

$\chi(G)$ The Chromatic Number of a Graph G

γ The exponent of the degree distribution $P(k) = k^{-\gamma}$

λ_i The i^{th} eigenvalue of the Adjacency matrix of a Graph G

Π_i The probability of attachment to node i

Π_i^c The constrained attachment probability of connection to a node i

Acronyms / Abbreviations

ANOVA Analysis of Variance of Two or More Data Samples

CMDB Configuration Management Data Base

CSV Comma Separated Values

FN False Negative

FP False Positive

FPR False Positive Rate

GDMO Guidelines for the Definition of Managed Objects

IEEE Institute of Electrical and Electronics Engineers

IP Internet Protocol

ITIL IT Infrastructure Library

ITZ Internet Topology Zoo

MTTD Mean Time to Detect

MTTR Mean Time to Resolve

OSI Open Systems Interconnection

RCA Root Cause Analysis

ROC Receiver Operating Characteristic

SDN Software Defined Networking

SVD Singular Value Decomposition

TN True Negative

TNR True Negative Rate

TNSM Transactions on Network and Service Management

TP True Positive

TPR True Positive Rate

Chapter 1

Introduction

1.1 Overview

1.1.1 Summary

In this thesis I will describe the work presented for publication in a number of academic journals. To provide sufficient background to the papers though, it is first necessary to explain the motivation for the research and the potential practical benefits it offers to the domain of Fault Management, Fault Localization in particular.

I begin with an overview of Fault Localization in Section 1.2.1, which describes the main commercial approaches to the use of event monitoring to identify faults in a monitored system, hopefully before impact is felt by end users. This is an important discipline, and most large operators of network and compute infrastructure have substantial investments in both the tools and manpower necessary to perform this task. In section 1.2.3 I describe how each of the approaches have fundamental limits in the rate at which they can process event data. Given that the vast majority of events monitored do not indicate a service impacting failure (for example in [54] this can be as little as 0.163% of events). Elimination of these spurious events would vastly increase the scalability of fault localization techniques.

The search for a method of elimination of these noisy events was the focus of my early research and the subject of the first two papers described in Chapter 4 and 5.

Before describing these papers, in Chapter 2, I introduce and summarize the necessary Graph Theory background used in the research. The majority of the analysis that I have undertaken relies upon concepts of Graph Entropy, which is a relatively esoteric mathematical construct, not often studied. I will attempt in my overview to provide some insight as to what it measures about the structure of a graph, and the networks we represent graphically.

During the course of the research aimed at understanding how network topology could be used to eliminate noisy events, I started to investigate the node degree distributions of the topology data. I noticed that in many cases the data did not follow the expected pattern of the most accepted model of network evolution. This became the focus of the last part of my research and resulted in the paper described in Chapter 6.

I consider the results obtained in all of the three papers presented in this thesis to be both novel and of practical use. In fact the vertex entropy technique is currently being implemented at Moogsoft Inc as part of its suite of event management products.

1.2 Fault Localization Background

1.2.1 Overview of Fault Localization

Operational Motivation

For the purposes of my research, I have focused on the important operational discipline of fault management as applied to large scale real world networks. The discipline concerns itself with the identification of failure conditions in components of the compute, application and network infrastructure of a business seeking to digitally support its commercial and regulatory activities. These failures can result in outages which impact consumers of these services. Using the latest verified census (statistics taken from [18]), it is estimated that $5.809 billions of dollars were transacted using digital infrastructure owned and operated in the private and public sector in 2014, with retail e-commerce growing at 14.3% annually. Clearly, conditions which render a digital service unavailable or unusably slow can have real and potentially existential affects upon a business.

The role of fault management is to identify these failures rapidly enough to permit the remediation of issues in the fastest possible way to minimize such impact. This benefit can often be quantized by companies as the cost of a minute of downtime, which can run into thousands and even millions of dollars, depending upon the business.

The typical process of Fault Management divides neatly into two phases, fault localization and fault remediation. The first phase is measured with the Mean Time to Detect (MTTD) metric, the second phase with the Mean Time to Resolve (MTTR) metric. Some of the basic challenges and strategies to overcome them are outlined in [78]. In the context of fault management this work gives a good example list (their Table 1), of the variety of source information that is typically available for fault localization, but to standardize terms in this work we introduce the following definition for a monitored event and an incident:

- *Event*: An event is typically a single log message or notification from an underlying monitoring system. We require that it has a timestamp, topology node identifier and description. It is not necessarily a notification of a fault condition, but fault conditions will send out at least one event.

- *Incident*: An incident is a support ticket raised as a result of some failure or service interruption. In the context of a management application, an incident is typically raised upon receiving an event that is deemed to be potentially indicative of such a failure, and each incident can be linked to a node in the network topology from which the event was received. Although not all incidents are indicative of a significant impact, they are an indication that the node has a fault condition that requires investigation.

Practically, each of these types of data object are comprised of a tuple of key value pairs, each key representing a fundamental property of the event or incident. Each source of monitored infrastructure will produce different event formats, and a common challenge before processing is to normalize this event data into a common format appropriate to the system being used to manage them. There is no absolute standard for the minimal description of an event (or incident), but it is expected that each event contains:

- *Timestamp*: Some record of when the event occurred, often measured in epoch seconds, commonly a UNIX timestamp.

- *Host*: A network name for the sender of the event, usually associated with the affected entity in the case of an event or incident. It is this field that relates the topology of the managed system to the events and incidents. Typically the management system will include in the host field the label from the underlying topology database and this will be also be stored in any incidents that are raised as a result of the event being received. It is common for this field to be a fully qualified domain name, IP address or similar.

- *Description*: A human readable description of the event.

- *Severity*: Some measure of severity of the event, in terms of its impact.

Typically events have many more attributes than those listed above, but any valid event must *at least* contain these, and the same is true of incidents. In the case of commercial systems an event may contain upwards of twenty core attributes (see for example [36]). Critically, though, through the 'Host' attribute it is possible to relate events, incidents and structural topology of a managed infrastructure

Herein lies the fundamental challenge. The number of events that need to be processed far outweighs the number of incidents typically experienced in a sizable network infrastructure. From practical experience with many very large environments from banks such as Royal Bank of Canada, to web scale companies such as Yahoo and GoDaddy. In industrial scale networks it is common for the monitoring systems to receive as many as 10^5 to 10^9 events daily, in which there maybe 10^2 to 10^4 incidents. Additionally in the literature such as Stearley *et al* [53] it is well understood that large scale infrastructure produces a large amount of event messages, most of which are not actionable. Fault localization is primarily the identification amongst the field of events, the subset that indicate the causal problems leading up to the incident. Should these events be properly localized the process of diagnosis and eventual remediation is significantly simplified.

Many different approaches can be taken to categorize an event as causal (and therefore interesting), or as background noise. In the following Section 1.2.2 we survey the main ones. What they all share in common are limitations to their ability to handle high event rates. A central goal of our research is to identify how the structure of a network can assist in lowering the event load. As well as identifying potential methods to do so, an interesting side-shoot of the research has been a novel way to understand how such networks evolve and grow over time.

1.2.2 Algorithmic Approaches

The collection of algorithms used to perform fault localization is commonly referred to as Root Cause Analysis (RCA). There are many approaches to doing this, but the fundamental input to the algorithm is the stream of monitored events, and the output is a list of correlated alerts (occasionally represented by one root cause alert). In addition to the received events, most of the algorithms require some description of the underlying system that is being monitored. In this section we briefly survey the most widespread approaches.

The Steinder-Sethi Categorization

An excellent review of fault localization techniques is presented in the 2004 paper by Steinder and Sethi [68], and contains a very useful categorization of the approaches to fault localization. We reproduce the relevant portion of that ontology in Figure 1.1.

This categorization of fault localization techniques attempts to group the various algorithmic approaches to fault localization into the broad classes of ‘AI’ techniques versus ‘Fault propagation models’. Inevitably in any such ontology, there is ambiguity between

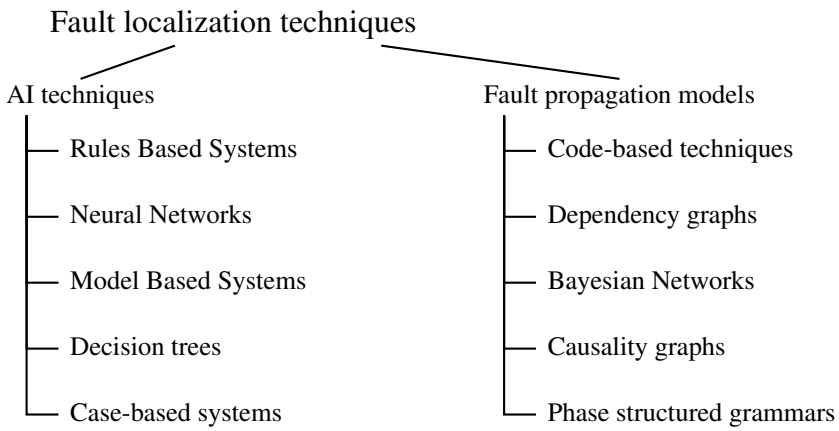


Fig. 1.1 Steinder - Sethi Ontology of Fault Localization Techniques

the categories of certain algorithms, for example the difference between decision trees and dependency graphs is a matter of opinion! For clarity a brief description of each category, and the relevance to our survey is as follows:

- **Rules Based Systems:** These algorithms apply a series of boolean if-then-else logical reductions to the set of events being processed with the result being the inference of a given root cause. We describe these in detail in Section 1.2.2 "Rules Based Systems" and Micromuse's Netcool is a primary example of a system deploying these algorithms.
- **Neural Networks:** Neural network based systems use a series of features of the events received and known set of incidents related to them, to train a neural net to recognize the presence of an underlying incident. These overlap the algorithms describe in Section 1.2.2 "Machine Learning Based Approaches" and are utilized in the products of Moogsoft.
- **Model Based Systems:** Model based systems rely upon the creation of a behavioral model of the managed system, including the events produced in response to modeled failure modes. These class of algorithms include those discussed in Section 1.2.2.
- **Decision Trees:** This class of algorithms attempts to codify the diagnostic process into a linear flow of tests for the presence of events, which is used to analyze for the presence or absence of events as symptoms or causes in the system. They are not considered in this survey as they are not in common usage due to the fact they require events to arrive in a precisely anticipated order and contain no unanticipated events [68].

- **Case-based systems:** Case based systems rely upon the creation of a knowledge base of past incidents and the events that were symptomatic and causal. Various techniques are then applied to the events that are current in a system to relate back to the case. We do not consider these algorithms here as they are not currently in widespread use due to scalability limitations [68].

- **Code-based techniques:** The code-based techniques use an efficient algorithm to match known event patterns for collections of incidents (described in Section 1.2.2 "Codebook Correlation" as problems) to deduce the most likely set of incidents to have caused the observed events. EMC Smarts is the most notable example of a commercial product to use this approach.

- **Dependency graphs/Bayesian Networks/Causality Graphs:** These algorithms use a weighted directed graph to model the manner in which failures propagate through the network by representing failures as nodes and the edges between them the probability of one failure causing another. In the case of a 'Dependency Graph', a one to one correspondence is then made between failures and events, and predicted events are compared against actual to deduce which failures are represented by the events received. Because of the one to one mapping of failures to events these approaches are not widely used, and not covered here, as in general a single failure can produce many events [68]. 'Bayesian Networks' and the closely related 'Causality Graphs' are a general form of the approach used in the codebook approach described in Section 1.2.2 "Codebook Correlation", which to the authors knowledge are the only widespread use of this class of algorithms. The generalization is the use of the network to model the links between failure nodes as conditional probabilities. Multiple nodes can connect to a single node, and the probabilities assigned to the weights of the directed links in the graph represent the conditional probability of the target of a link in the graph occurring if the source node exists. In this way multiple events as symptoms of a given failure can be modeled, and the algorithms proceed by attempting to decide the most likely failures present from the events being analyzed.

- **Phase structured grammars:** This class of algorithms utilize a context free system of grammar to model the complex interdependencies between components in a system. It accomplishes this by decomposing a dependency graph into a set of paths that represent how failures propagate. Linear programming techniques then attempt to match these paths to the largest subset of analyzed events to determine root cause. They are not in common usage and I do not survey them here as the approach can be considered to be

a subset of the codebook approach described in Section 1.2.2 "Codebook Correlation" [68].

A key missing category, which has emerged since Steinder published the ontology, are systems that use data driven algorithms and machine learning to localize faults. This is a central focus of companies such as Moogsoft [48], and I will spend some time in Section 1.2.2 describing them.

A shared characteristic of each of these approaches is the use of advanced logic to inspect the events that come from the source systems to determine the potential impact of the event. The additional processing that this logic requires effectively throttles the maximum rate at which events can be processed. It is a fact of modern infrastructures that the event rates that are operationally required to deal with increase inexorably as the scale of the networks increase. In essence the objective of fault localization is to correlate a very small subset of the received events with service impacting outages. This *'event correlation'* is a term more common in industry and is used to represent the entire family of fault localization techniques.

Rules Based Systems

Among the first attempts to solve the event correlation problem was the use of rules to model the behavior of the managed systems. In essence these systems operate as a finite state machine that models problems as states of the system. Transitions between states are triggered when a filter matches an event being processed by the system. If the filter matches the system transitions state, if it does not match the state remains unchanged. In principle every possible transition is considered for every event arising. A very early example of such a commercial system was NetExpert, originally developed by OSI Inc [51].

A very simple example of such a system is depicted as a decision tree in Figure 1.2. The basic operation relies upon the rules being able to make processing decisions by the application of a suitable logical language on the data contained in the event, or by referencing a secondary database of information (typically a Configuration Management Database or CMDB). In the example depicted in Figure 1.2, the state machine operated by the rules based system is set up to match three different types of events A, B and C. The precise nature of how these matches are determined depends upon the system deployed, but in essence it is usually expressed as a set of conditions on the attributes of the events, such as source host and textual description. In the depicted state machine, the matching of events progresses the system through 3 states, any of which may be associated with critical conditions and cause the initiation of some operational intervention (the paging of a support person for example). The

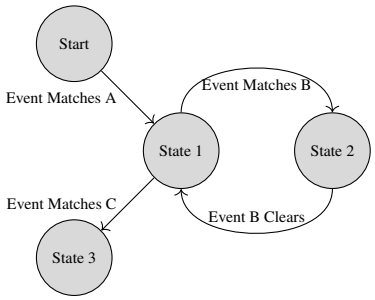


Fig. 1.2 Example of a Basic Rules Based System

transition from State 2 back to State 1 is triggered when Event 'B' is cleared, that is the event condition is either manually or automatically closed. This type of logical flow is common in fault management systems and is to date the most widespread type of fault localization methodology. The usual language to express the decision logic is a combination of boolean expressions operating on attributes of the events. We reproduce in Table 1.1 an abbreviated subset of the common core components of an event message. There is no absolute standard of what is necessary in an event, though many standards bodies have attempted to produce such a reference. The most widespread set of standards that rules based systems operate to is contained in the IT Infrastructure Library (ITIL), a set of manuals produced by the British Government [52].

Label	Typical Contents	Data Type
Host	Fully Qualified Domain Name of sending host	Textual
Severity	Notion of Impact, Critical->Clear	Enumerated
Timestamp	Recorded UTC time that event was originated	Numeric, Epoch Seconds
Description	Human readable descriptive message	Textual
Impacted Service	Indication of dependent services	Variable/Textual

Table 1.1 Description of Type and Content of Typical Event Attributes

Using the attributes in Table 1.1 an example for "Rule 1" could be of the form $((Host =$ "Important Host"$) \wedge (Severity >$ "Minor"$)) \implies$ "State 1". In practice implementations of these types of systems involve *thousands* of such expressions, often built over many years. As every event that arrives at the system must be evaluated through every transition between states, the processing load as scale increases in terms of event rates and complexity of the state engine can become prohibitive. Although the transition filters can be very simple, in general it is not possible to make every transition independent from every other transition, which prohibits an aggressive parallelization strategy for the evaluation of the finite state machine. As a result, the cost, and even effectiveness, of this approach is questionable and is

only widespread in older, legacy environments where event rates are low and the underlying infrastructure is fixed. Precise benchmarks on the performance of rules based systems tend to be closely guarded, but in papers such as Gardner *et al* [32], and in Steinder *et al* [68] and Kilger *et al* [39] the highest rates for these systems are measured in thousands per second. In more modern infrastructures this limitation is prohibitive, and the use of rules based systems in large scale virtualized infrastructures is becoming less common.

Codebook Correlation

The codebook approach was pioneered in the early 1990s, as is described in Kliger *et al* [39], as an attempt to overcome the drawbacks of the rules based systems described in Section 1.2.2. The approach draws upon methods of probabilistic inference on a causality graph. A causality graph is a representation of cause ('Problems', which equate to incidents or root causes) and effect (Symptoms, which equate to events). Once this graph is established, the assumption is that every event received by the monitoring system is either noise, or a direct result of a problem having occurred creating the symptom as represented by the event. As different problems can emit the same events, the core operation of the algorithm is to distinguish the most likely set of problems that must have existed to result in the observed symptoms. To further improve the predictive power of the model the causality graph that is built can have weighted edges to indicate the probability that a given problem will result in a connected symptom.

The implementation of the algorithm translates the causality graph into a table, which is essentially a weighted adjacency matrix for the graph, but it is typically non-square as the number of problems and symptoms are not normally equal.

Once the codebook is calculated from the causality graph, it is used to compare against the set of events that are received. The events are represented by a feature vector, so for example if S1 and S3 were recorded the event vector is $\mathbf{E} = (1,0,1)$. In this instance the matching of this vector to a single cause is ambiguous and herein lies the principal drawback of codebook. In the analysis detailed in [39], the effectiveness is measured by the 'radius' of the codebook, which is defined as half the minimal Hamming distance between codes (columns in the Table 1.2). In the description of their algorithm Kliger *et al* describe a 'generalized' Hamming distance as being the sum of the integer differences between each of the vectors. In this way the Hamming distance between two vectors $\mathbf{A}$ and $\mathbf{B}$, $d_H(\mathbf{A},\mathbf{B}) = \sum_i |a_i - b_i|$, where a_i and b_i are the components of $\mathbf{A}$ and $\mathbf{B}$. For example if $\mathbf{A} = (1,0,1)$ and $\mathbf{B} = (1,1,1)$, the Hamming distance, $d_H(\mathbf{A},\mathbf{B}) = 1$. It is asserted that providing that the radius is greater than 0.5 then the codebook is effective at distinguishing between candidate problems for any given

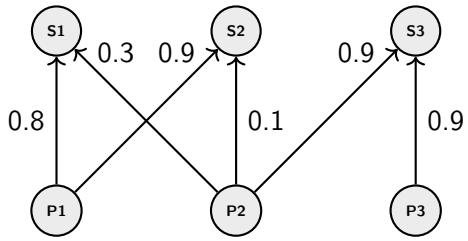


Fig. 1.3 A Simple Codebook Causality Graph

event vector. No proof is offered of this assertion, but the argument is plausible as Hamming distances of less than 1 would not be obtainable, as it would indicate no differing symptoms.

Symptom	P1	P2	P3
S1	0.8	0.3	0.0
S2	0.9	0.1	0.0
S3	0.0	0.9	0.9

Table 1.2 Codebook for Causality Graph in Figure 1.3

The generation of codebooks is a significant drawback of the approach, as detailed analysis of the cause and effect in the monitored system must be undertaken. The most classical use of event correlation is the so-called 'downstream suppression' problem. A common approach to network monitoring involves availability checks being undertaken on a periodic basis. In the case of an IP network, this is often accomplished by sending an ICMP PING packet to each of the nodes in the network [71]. If the node is operational and a path exists to the node, an echo will be received and the node will be deemed to be functional. There are many ways in which an echo may not be received however, that range from the node not being operational, to a communications problems on the path from the monitoring host to the monitored node. They may arise from intervening nodes being in a failure state, or, other more complex issues with the configuration of the network. If a node close to the monitoring host fails, this may cause spurious 'ping fail' events to be received from all nodes that are 'downstream' from the failed node and in a large network this can produce a significant amount of false alerts. Typically, one imagines that behind a switch are a collection of important servers that are being monitored by such a periodic poll from a point in the network. Should this switch malfunction events would be expected to be received from all of the devices. We illustrate this example in Figure 1.4, and the corresponding codebook in Table 1.3. In the table we present the symptoms S1,..,S4 as events received from the management system when there are poll failures of the servers (S1,S2,S3), and a

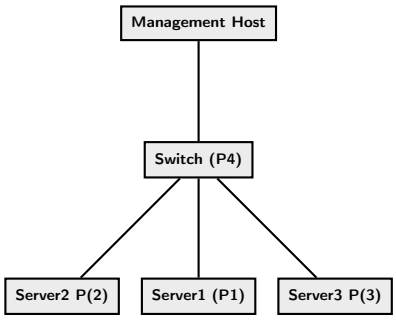


Fig. 1.4 A Simple Downstream Topology

Symptom	P1	P2	P3	P4	P(1,...,4)
S1	1	0	0	1	1
S2	0	1	0	1	1
S3	0	0	1	1	1
S4	1	1	1	1	1

Table 1.3 Codebook for Figure 1.4

poll failure of the switch (S4). The corresponding root causes, or problems P1,...,P4 are the failure of a single server (P1,P2,P3), and a failure in the switch (P4). The problem P(1,...,4) is the scenario when all of the devices have failed. The radius of this very simple cookbook, which is nevertheless a common management scenario, is zero, and the algorithm cannot distinguish between the failure of the switch and the failure of the switch and all connected devices. This could occur, for example, in the event of a power failure in a data center. In fact, as simple as it may seem, this example is one of the reasons that event rates are so high, as most events are consequential rather than causal. This very low ratio of events from true causes, versus events from either consequentially affected nodes or simply noise, is one of the central problems in fault localization. The downstream example above is a very basic example how one failure can cause the emission of events from functionally operable nodes that are simply no longer visible from the management server. In the paper of Kliger *et al* [39] and also in the work of Stearley *et al* [53, 66] this low quality of event information is described. In particular, in a study of 178 million syslog messages in [53], nearly 79% of the messages were not indicative of a real problem.

Many attempts have been made to improve upon the basic codebook technique, both in codebook generation and improvements to the causality graph approach (for example see ([87] and [8]). These range from reducing the complexity of generating the cookbook in the case of [8], and suggested improvements to eliminating the effect of noise [87]. The fundamental drawback remains which is the reliance upon a well understand failure propagation model for the system, which may not be available or, at best, complex and expensive to build and maintain. In any case, this approach, and other similar ones based upon probabilistic causality graphs have also been proven to be *NP-Hard* [20], which in practice means that they are not suitable for use in the large scale environments that I have studied during my research.

Topology Driven Active Object Model

At the same time as the codebook approach was being developed an alternative approach was developed using a topology driven active model. This relied upon an accurate topology of the network being obtainable that reflected the detailed interconnectivity of components of the network, such as is depicted in Figure 1.4.

The Active Object Model was built using an extension of the ITU-T, GDMO (Guidelines for the Definition of Managed Objects, see [37]) standard, part of the Open Systems Interconnection (OSI) suite of standards (see [89]). The extensions proposed allowed specific object classes to be imbued with management behavior, triggered by the addition of specific events to instances of these classes. A specific use of this behavior was to send messages to other 'Active Objects' when an event was received indicating that the represented system was in a failure mode. This message could in turn be propagated to other connected devices, in particular in a direction that is 'downstream' from the monitoring system. In this way the system could handle the downstream suppression use case very effectively.

The correct operation of such a system is dependent upon an accurate topology. The acquisition of this topology in a large network is not straightforward and the limitations of using commonly known approaches such as `traceroute` are well understood (as explained in [84] and [40]), but fundamentally arise from the fact that Internet Protocol (IP) operates at a layer in the network which is independent of physical connections and corresponding devices such as ethernet switches. To overcome these limitations a multi-layer discovery system was developed inspired by the 'Fremont' discovery prototype developed at the University of Colorado by Wood *et al* [85].

Once the discovery was completed, for each discovered device an appropriate active object was instantiated, and the collection of Active Objects was stored in a memory resident database for efficiency. Re-discovery of the network was continuous to capture changes in configurations, but once discovered the system was capable of monitoring and performing fault localization in real time and at scales of many hundreds to thousands of events per second, somewhat in excess of the capability of the codebook approach. The basic algorithm is described in Algorithm 1. This algorithm uses a model of all paths through a managed network to determine which nodes are downstream of any given node in the network. As each event arrives at the system a recursive walk through the network is undertaken from the node referenced by the event to all downstream nodes of it, closing any non-causal events that are consequential of the analyzed event. The system supervises the recursion to prevent looping through the topology. The suppression is enabled by each node having a class definition that identifies which types of events are suppressible if a given event exists upstream from it.

Nevertheless, accuracy was highly dependent upon the topology being complete and up to date. In the event that this is not the case, incorrect results would be obtained, most dangerously false negatives where a real problem is suppressed because an 'upstream' device in the topology has a fault, but is no longer 'upstream' in the monitored network. The emergence of fast changing topologies such as Asynchronous Transfer Mode (ATM) and Software Defined Networking (SDN), makes the application of this approach impractical.

Algorithm 1 Downstream Suppression Algorithm

```
1:  procedure SUPPRESSDOWNSTREAM(Node n, Node managementHost, Event e)
2:      downstreamNodes ← nodes downstream of n path initiating at managementHost
3:      activeObjectClass ← class of n
4:      eventsInNode ← all events on node n of class activeObjectClass
5:      for all Event event in eventsInNode do
6:          ▷ Check if each event for this node is suppressible by the event e and close
7:          if event closable by e in activeObjectClass
8:          when downstream of managementHost then
9:              CLOSEEVENT(event)
10:     ▷ Iterate over all downstream nodes, recursively calling procedure
11:     for all Node downstream in downstreamNodes do
12:         SUPPRESSDOWNSTREAM(downstream,managementHost,e)
        return
```

Machine Learning Based Approaches

Moogsoft is a pioneer in the application of Machine Learning techniques to the detection and isolation of faults. The earliest example of data science being applied to the problem is the work of Stearley *et al* described in [66],[67], and [54]. Their approach makes use of Shannon Informational Entropy [62] to attach a metric to each event log received from an array of managed super computers. The system that was built, 'Sisyphus', was presented and described at a number of conferences but never gained widespread use.

In part this was because the entropy metric merely served to rank the importance of an event log. In modern contexts the rate at which these log messages may be produced (often measured in tens of thousands of messages a second) would produce too many high importance events to be of use. As a pre-conditioning metric this is a good starting point to create a flow of high quality data to feature detection algorithms.

It is precisely this approach which motivated the research that resulted in the papers presented in this thesis, and substantially forms part of the product offering of Moogsoft [48]. The Moogsoft suite of software conducts entropy analysis similar to that described by

Stearley, and then pipelines the events into a suite of 'Sigalisers' that seek to group together log messages when they are in some way anomalous and believed to be related to an incident that requires human intervention. A Sigaliser is a module of the Moogsoft software that is responsible for surfacing clusters of alerts that are deemed to be causal in some sense. In this context causality means that they are caused by an underlying incident that is actionable, and each alert in the generated cluster is a valid symptom of the underlying cause.The Sigalisers use a variety of data driven algorithms including:

- *Time Based Matrix Factorization*: This approach, based upon a Non-Negative Matrix Factorization technique, initialized using a Singular Value Decomposition (SVD) initialization (see [15]), attempts to group together events that have significant correlation in terms of the pattern of temporal occurrence.
- *Language Based Similarity*: This approach clusters events which share significant textual similarity across a selection of attributes. This is done with an SVD initialized k-means approach (see for example [86]).
- *Structural Proximity*: This approach clusters events when they are from nodes that are a small distance in terms of network links, or hops, from each other.

Each of these approaches has been demonstrated to have great practical benefit in terms of both the precision of the clusters generated and the recall (as measured by those incidents detected by monitoring as a fraction of all known incidents).

A significant practical advantage of the machine learning based approach is the independence the algorithms have from a pre-conceived model of failure modes of the systems being monitored. In modern infrastructures, in which configuration change is constant and often instant, this is a significant impediment to techniques such as codebook, or rules based fault localization. The penalty for this resilience is the requirement of significant computational power required to conduct the algorithmic analysis of the events.

1.2.3 Known Limitations of Fault Localization

In essence all of these approaches to fault localization suffer from some or all of the following problems:

- Scalability: As all known approaches involve the individual inspection of each event, the entire system will have a maximum throughput it can consume. These limits range from very low (for example in Table 1 of [17] at best 30ms is taken to process each

rule and therefore event which implies a throughput of 34 events per second), to a few thousand a second (these are claims typically made by commercial systems [36], [25]). Anticipated loads in commercial networks may easily exceed these.

- Accuracy: As demonstrated in Section 1.2.2 even sophisticated algorithms are vulnerable to the inability to distinguish between root causes with the available evidence. Ultimately this drives a search for ever more complex approaches, inevitably placing further constraints on scalability.

- Tolerance to Change: With the exception of machine learning, all of the known fault localization techniques require a detailed model of the monitored system to be able to localize faults. In older implementations this was not a significant obstacle as the typical data center technologies in common use before the advent of Software Defined Networking (SDN), cloud computing and related innovations had change cycles that permitted the editing of these models.

The need for accurate fault localization techniques is a critical operational requirement of datacenter technology, and it is vital that techniques and methods are developed to mitigate these issues. This is the primary original motivation of the research presented in this thesis.

The Need For Event Rate Suppression

A common thread of the issues described in the previous section is the need to contain the rate of events that fault localization approaches need to process. Without some form of event rate suppression it is anticipated that most model based techniques will fail to be of practical use as networks scale and adopt modern technologies. It is not anticipated that event rates will do anything but continue to increase, and so the search for methodologies to cut down event rates is of considerable commercial interest.

The techniques described in the first two papers of this thesis (Chapters 4 and 5), offer the possibility of a way to process network configuration quickly to permit the discarding of up to 65% of events whilst only sacrificing 20% of incidents. Although missing 20% of incidents may seem operationally dangerous, it is important to stress that processing 65% of events discarded may actually result in many fault localization systems missing substantially all of the incidents, due to the scale limitations mentioned above

1.2.4 Summary of Research Activities

During the course of my studies, the research activities followed the program outlined in Figure 1.5, resulting in the publication of three papers, attendance at a number of academic interest conferences, submission of three papers which were not published, and seminar presentations. The principal activities include:

- Submission of a poster paper to SIGCOMM 2015, which was not accepted.
- Submission of a conference paper to IEEE NOMS, AnNet workshop in Istanbul, April 2016 which was published [75].
- Attendance of MoN14 in September of 2015 at Oriel College Oxford.
- Presentation of paper to the AnNet workshop, at which I won the best paper award.
- Submission of a paper to the IEEE Infocom in Atlanta May 2017, which was not accepted.
- Submission of a paper to the International Journal of Network Management, which we subsequently withdrew in favor of a submission to the Transactions on Network and Service Management.
- Delivered a work in Progress Seminar at the University of Sussex in June 2016.
- Submitted a talk to MoN15 at the University of Bath, which was accepted [72], which also resulted in a collaboration with Prof Jonathan Dawes.
- Submitted a full journal paper to Physics Review E on Constrained Attachment, arXiv preprint here [73],which was rejected by one reviewer after two rounds.
- As the Constrained Attachment paper was accepted by one of the reviewers, the paper was resubmitted to European Physical Journal B, and is currently being reviewed following resubmission after one round of comments.
- Invitation to join the Technical Program Committee for AnNet 2017 to be held in Lisbon in May 2017.
- Submission of a paper to the IEEE Transactions on Network and Service Management journal, which was accepted for publication [74].

Initially my research was focused solely on the availability of vertex entropy measures of node importance for use as an automated event filtering mechanism for fault management. This was motivated by the access I have to a number of interesting real world data sets, including months of event, incident and topology data from some large customers of Moogsoft. A current, and important problem when deploying event management systems is the control of event rate, and the use of available topology information to do so, seems a logical step. This was the subject of the first three attempts at publication, but during those studies, and as a result of exposure to theories of dynamic graph evolution, particularly the standard scale free model of Barabási *et al* at the MoN14 conference, I became interested in dynamic models of network growth [2]. In particular deviations from the predicted power law degree distributions were noticeable in all of the data sets that I had access to, which lead to the second major focus of my research, the effect of connectivity limitations in communications networks on dynamic network evolution. I developed a new model of network evolution which I presented at MoN15, which was well received, and indeed initially intended to publish a paper solely on that model. The final focus of my research was motivated by a potential link between vertex entropy and dynamic network evolution. After experimenting with a toy model of the statistical mechanics of network evolution from a paper by Newman and Park [56], I developed my own model using vertex entropy. Ultimately this lead to the final publication submission to the Physical Review E, and simultaneous open publication on arXiv. Although the paper was rejected by Physical Review E, one of the reviewers accepted the paper and was positive about the contribution. After consultation with my co-authors, the paper has been resubmitted to European Physical Journal B, and is currently going through its second round of review.

The future direction of my research is very exciting, as the entropic model of network evolution opens up the possibilities of a different approach to analyzing the robustness and aging of many networks, including but not limited to communications networks. I have developed a number of very good collaborative relationships beyond my supervisors, with academics at Sussex (Dr István Kiss and Luc Berthouze, the former collaborated on the paper in Chapter 6), University of Bath (Professor Jonathan Dawes was a co-author on the paper in Chapter 6), Queen Mary University of London (Dr Richard Clegg, organizer of MoN) and Oxford University (Dr Justin Coon). I am keen to maintain my involvement in the network science research community after the completion of my doctoral studies.

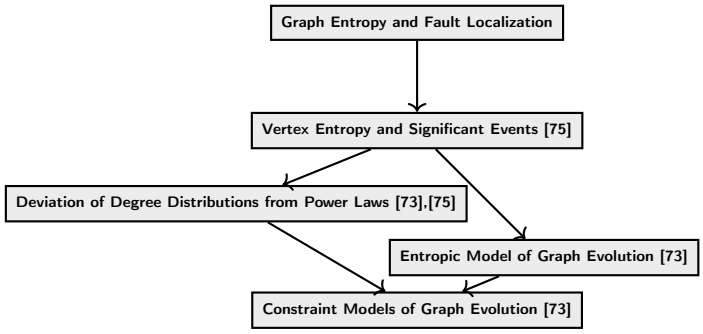


Fig. 1.5 The Journey From Fault Localization to Graph Evolution

Chapter 2

Graph Theory Themes

2.1 Graph Theory Essentials

2.1.1 Basic Definitions

In this section I will describe the essential mathematical concepts of Graph Theory, that are referenced in the published work. It is not intended to be a comprehensive treatment and I lean heavily on the standard texts of [13], [14] by Bollobás, and in the section on graph evolution the recent book by Barabási, [6].

A graph, in the mathematical sense, is a collection of two sets. The first, commonly written as V is the set of nodes or vertices, and the second E the set of edges $e_{i,j}$ that connect two vertices v_i and v_j. In fact, more rigorously, $E \subseteq V \times V$, which, for a simple undirected graph that we define below in Definition 1 and 2, can have maximum size $\frac{1}{2}n(n-1)$, where $n = |V|$. We refer to the graph by enumerating the two sets as $G(V,E)$.

In the most general case, the edges in a graph can be oriented, and the graph is referred to as a 'directed graph'. In a directed graph, for a given vertex $v_i \in V$, we refer to its degree as the cardinality of the subsets of E consisting of all edges that begin at v_i and those that end at v_i. For the first set of edges originating at v_i, we count the 'out-degree' of v_i and for the set of edges terminating at v_i the 'in-degree' of v_i. The total degree, written k_i is the sum of in and out degrees.

In all of our research we focus upon undirected, simple, connected graphs, which we define, following the terminology of Bollobas [13] as:

Definition 1. *A graph $G(V,E)$ is said to be undirected if $\forall\ e_{i,j} \in E$, $e_{i,j}$ is indistinguishable from $e_{j,i}$. Note that in this case the in and out degrees are the same as the total degree of every vertex in the graph.*

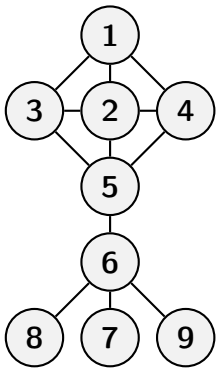


Fig. 2.1 An Example Graph

Definition 2. *A undirected graph $G(V,E)$ is said to be simple if it contains no loops (edges that start and end at the same vertex), and for any pair of vertices v_i, v_j, there exists only one edge $e_{i,j} \in E$ or $e_{j,i} \in E$.*

Definition 3. *A path of length n, $P(n)$ is a sequence of vertices $v_i \in V, i = 1 \ldots n$, such that each vertex $v_i \in P$ has an edge $e_{i,i+1} \in E$, for $i \in [1, n-1]$. A connected graph is a graph $G(V,E)$, such that for any pair of vertices $x, y \in V$, there exists a path $P(n)$ with $v_1 = x, v_n = y$.*

In the case of an undirected graph there is no distinction between in-degree and out-degree, and the degree of a vertex $v_i \in V$ is simply the number of edges that are incident to the vertex in the graph.

In Figure 2.1 we draw a simple graph that contains many of the features that would be encountered in a communications network. Nodes 1 to 5 form part of a highly meshed core, and 6 to 9 a hub and spoke topology, often seen in the access portion of a network. In terms of the two sets, we have $V = \{1,2,3,4,5,6,7,8,9\}$ and $E = \{e_{1,2}, e_{1,3}, e_{1,4}, e_{2,3}, e_{2,4}, e_{2,5}, e_{3,5}, e_{4,5}, e_{5,6}, e_{6,7}, e_{6,8}, e_{6,9}\}$.

A graph is uniquely defined by listing all of its edges, but it is possible to identify interesting symmetries in the graph by relabelling vertices. Following the treatment in [13], two graphs are said to be *isomorphic* if under a re-labelling of vertices the set of edges is identical, so that only edges that existed between any two vertices before re-labelling, exist after the re-labelling. Formally two graphs $G(V,E)$ and $H(V',E')$ are said to be isomorphic if there is a correspondence between the vertex sets which is adjacency preserving. Recall that a bijection is a map between two sets A, B of identical size, $f : A \to B$, which maps every member of A to a unique and distinct member of B, so that every member of B can be mapped back to a unique and distinct member of A. Using this definition, two graphs $G(V,E)$ and $H(V'E')$ are isomorphic if there exists a bijection $f : V \to V'$ such that an edge between

(a) K_4 (b) S_4 (c) P_4 (d) C_4

Fig. 2.2 Special Graphs of Order Four

vertices $x, y \in V$ exists, if and only if, there is a corresponding edge between $f(x), f(y) \in V'$. We can also define permutations of the vertex set V of $G(V,E)$, $\sigma(v)$, for each vertex $v \in V$, which can be represented as a relabelling of the vertices of the graph. These permutations allow us to define an *automorphism* of the graph, which is simply a permutation of the vertices, such that the graph with permuted vertices is isomorphic to the original graph. The set of automorphisms form a group that is used in the original definition of structural graph entropy, which is important in our later work on vertex entropy.

Some Special Graphs

In the theoretical treatment in the papers presented in Chapters 4 and 5, all of the graphs are simple, undirected graphs that are fully connected, and reference to a number of special graphs are made as they are directly relevant to communications networks. We reproduce the definitions from the papers as follows:

- **The Complete Graph (K_n)**: This graph is formed from a set of n vertices, maximally connected.
- **The Star Graph on n Vertices (S_n)**: This graph has one vertex v which is connected to all other vertices, with no other edges in the graph.
- **The Path on n Vertices (P_n)**: This graph is a simple chain of n vertices, connected by a single edge with no loops. The path has a single start node v_1 and end node v_n.
- **The Cycle on n Vertices (C_n)**: This graph is a special case of P_n such that $v_1 = v_n$; each node has degree 2.

In Figure 2.2 we present simple examples of these special graphs with $n = 4$.

2.1.2 Adjacency, Coloring and Clustering

Of particular importance when quantifying the structure of a graph are Adjacency, Coloring and Clustering. We take this opportunity to introduce these concepts that are somewhat foundational to the notions of graph entropy which motivates all of our research.

For each node $v_i \in V$, for a graph $G(V,E)$, the set of nodes that are neighbors of v_i form a local neighborhood of the vertex. This local neighborhood and the edges between these nodes, describe the local topology of the graph to the node. We can summarize this for a whole graph by the definition of the Adjacency matrix of a graph as follows:

Definition 4. *The Adjacency matrix* $\mathbf{A}_{ij} = 1$ *if* $\exists e_{i,j} \in E$, *and* $\mathbf{A}_{ij} = 0$ *if* $\nexists e_{i,j} \in E$, *or* $i = j$.

Related to the Adjacency matrix are the Degree and Laplacian matrix, which for convenience we define here as:

Definition 5. *The Degree matrix of a graph G is defined as* $\mathbf{D}_{ii} = k_i$ *and* $\mathbf{D}_{ij} = 0$ *if* $i \neq j$.

Definition 6. *The Laplacian matrix of a graph G is defined as* $\mathbf{L}_{ij} = \mathbf{D}_{ij} - \mathbf{A}_{ij}$.

There is a profound relationship between graphs and matrices, and for a complete survey please consult [4]. In particular the eigenvalues and powers of this matrix yield powerful clues as to the structure of a graph, For example $\mathbf{A}^n_{ij}$ computes for each value of i, j the number of paths of length n between i and j (for a proof see [38]).

The concept of adjacency leads naturally to the coloring of a graph. This analysis decomposes a graph into sets of vertices that are not neighbors. The term coloring has its origin in the famous map coloring problem where one seeks to color countries in a map of the world in such a way as no two countries of the same color share a boundary. In a directly analogous way, you can assign a color to each vertex of a graph so that no two vertices that share an edge have the same color. This leads us to the definition, following the treatment in [13], of a *Chromatic Class* of vertices as follows:

Definition 7. *A Chromatic Class of a Graph* $G(V,E)$ *is any set of vertices* $X \subset V$, *such that no two vertices* $v_i, v_j \in X$ *have a corresponding edge* $e_{i,j} \in E$. *That is it is a collection of non-adjacent vertices of* $G(V,E)$. *In the context of a coloring, each vertex* $v_i \in X$ *can be labeled with the same 'color'.*

In general there are many possible colorings, and therefore chromatic classes of a graph, but the minimum number of colors required to completely partition the graph into a set of chromatic classes is termed the 'Chromatic Number' of the graph and is denoted as $\chi(G)$.

The Adjacency matrix defines a natural set of eigenvalues and eigenvectors stated in terms of the eigen problem equation,

$$\begin{aligned} \mathbf{A}\mathbf{v} &= \lambda\mathbf{v}\,, \\ \text{which only has non zero solutions for } \mathbf{v} \text{ when,} & \\ \det(\mathbf{A}-\lambda\mathbf{I}) &= 0\,, \end{aligned} \tag{2.1}$$

where $\mathbf{v}$ is an eigenvector, $\mathbf{I}$ is the identity matrix, λ is the eigenvalue, and $\det()$ is the determinant of a matrix. This equation has a spectrum of solutions for $\mathbf{v}$ and λ, and this collection of eigenvalues λ_i contain important information about the structure of a graph, In particular, in our exploration of the extrema of entropies for graphs, Wilf *et al* proved that the Chromatic number $\chi(G)$ of a graph is bounded by $\chi(G) \leq 1 + k_{max} \leq 1 + \lambda_{max}$, where k_{max} is the maximum degree of a node of the graph and λ_{max} is the largest eigenvalue of the Adjacency matrix (see [83]).

Another important property of a node in a graph is the degree to which it is clustered. This is usually defined as the number of edges that the neighborhood of a node has compared to a fully connected graph. The maximum number of unique edges that an undirected, simple graph of n nodes may have is $\frac{1}{2}n(n-1)$. Given a node of degree k has k neighbors, if e is the number of edges between neighbors of a node v, but not connecting to v, the normal clustering coefficient $C(v)$ is defined as:

$$C(v) = \frac{2e}{k(k-1)}\,. \tag{2.2}$$

Clustering can also be equivalently defined in terms of the number of triangles, that is fully connected subgraphs of three nodes, in a way which is more general than the one given above as it is valid for directed as well as undirected graphs. If λ_v is the number of triangles that contain the node v, and τ_v the number of collections of 3 nodes in which v is connected to the other two, then we may define the clustering coefficient as:

$$C(v) = \frac{\lambda_v}{\tau_v}\,. \tag{2.3}$$

In the theoretical analysis presented in the papers, use is made of the j-Sphere, S_i^j, centered at the i^{th} node, introduced by Dehmer in [23]. Dehmer's original definition relied upon subsets of vertices of a fixed distance from a given vertex v_i. where distance $d(v_i, v_j)$ is the shortest distance between distinct vertices v_i and v_j (i.e. $i \neq j$). This definition excluded

the vertex v_i, and other interior nodes for $j \geq 1$, but this introduces zeroes for special graphs such as S_n, when calculating the traditional clustering coefficient. This is because for a star graph, no edges exist between the neighbors of the central node, or triangles containing the central node. In our analysis we seek to divide by the clustering coefficient, which would introduce infinities. The slight modification described in Equation (2.6) below, removes these zeroes for any connected graph. Our analysis extends the definition of a j-sphere as follows:

Definition 8. *For a node $v_i \in V$, we define, for $j \geq 1$, the 'j-sphere' centered on v_i as:*

$$S_i^j = \{v_k \in V | d(v_i, v_k) \leq j\} \tag{2.4}$$

and for convenience when we define the clustering coefficient in Equation (2.6), the related 'j-edges' E_i^j as

$$E_i^j = \{e_{k,l} \in E | v_k \in S_i^j \text{ and } v_l \in S_i^j\} \tag{2.5}$$

Using the extended version of the j-sphere in Equation (2.4) the generalized clustering coefficient of a j-Sphere centered at i is defined as:

$$\begin{aligned} C_i^j &= \frac{2|E_i^j|}{|S_i^j|(|S_i^j|-1)}, \\ \text{and for } j = 1, \text{ as } |S_i^j| &= k_i + 1, \\ C_i^1 &= \frac{2|E_i^1|}{k_i(k_i+1)}, \\ \text{where } |E_i^j| &\text{ is the cardinality of the } j\text{-edge set.} \end{aligned} \tag{2.6}$$

These definitions are used in the treatment of vertex entropy in the papers of Chapters 4 and 5.

2.2 Graph Entropy

The notion of the entropy of a graph builds upon the basic concepts of Shannon Entropy first introduced in 1948 by Claude Shannon [62]. This argues by analogy with Statistical Thermodynamics that it is possible to define the entropy of a series of signals in terms of the amount of choice, or uncertainty in choosing a given signal. Say for example that a source of signals emits an alphabet of n possible signals $X = \{x_1, x_2, \ldots x_n\}$, with a probability of $P_i, i \in 1, 2, \ldots, n$ in a given fixed time period. It is then possible to define the entropy $H(X)$ as follows:

Definition 9. *For a system or alphabet of signals $X = \{x_1, x_2, \ldots x_n\}$, where each signal x_i is emitted with probability $P_i, i \in 1, 2, \ldots, n$, the entropy $H(X)$ is defined as*

$$H(X) = -\sum_{i=1}^{n} P_i \log_2 P_i \tag{2.7}$$

The entropy of this alphabet of signals is maximized when the probability of emission of each of the possible signals is uniformly distributed, which has the effect of minimizing the relative information conveyed by the arrival of a particular signal. It is possible to prove this assertion using the method of Lagrange, closely following the proof of Theorem 1 in the paper presented in Chapter 5. It is argued in the original work of Shannon [62] that the more uncertainty in which signals will be received, the less information is contained in them. This is the case when the probability of each signal in the alphabet is uniform and little information is conveyed by the arrival of any one signal. This situation can be compared with the sequence of characters in a spoken language, such as English, where each character has a very different occurrence pattern, and uncommon characters such as 'q' can very rarely be followed by anything other than 'u'. In this sense 'q' carries a lot of information, and conversely the entropy of the alphabet of the English language is low. This basic definition has been extended to characterize the entropy of a graph in a number of ways, which are outlined below.

2.2.1 Körner or Structural Entropy

In [41], and beautifully explained in [63], János Körner introduced the concept of the entropy of a graph in terms of a modified version of Shannon's original argument. Considering the alphabet X, as defined above, imagine that not all of the signals are distinguishable. A graph can be constructed by mapping to the vertex set V each of the signals in the alphabet, so that $v_i \in V$ equates to x_i and naturally associated with each vertex is a probability of emission of a signal $P(v_i) = P_i$. Now, each of the vertices are connected with an edge $e_{i,j} \in E$, if and only if the two signals x_i, x_j are distinguishable. The original definition given by Körner is highly technical and defined in terms of the independent sets and automorphism groups of the graph, where an independent set is equivalent to a chromatic class, that is a collection of vertices that are not adjacent. An alternative, more consumable definition is given in Section 1.5 of [50], which relies upon the notion of conditional entropy. Before reproducing that definition let us first define what is meant by conditional entropy, following the treatment in 'Information Theory' by J. Stone [69].

Definition 10. *Consider two random variables X, and Y, which can take the following values from a set of n signals $X = \{x_1, x_2, \ldots x_n\}$, and m signals $Y = \{y_1, y_2, \ldots y_m\}$. These variables are chosen according to the probability distributions $P(x_i), i \in [1,n]$, $P(y_j)\, j \in [1,m]$, and joint probability distribution $P(x_i, y_j), i \in [1,n], j \in [1,m]$. We then define the conditional entropy $H(X|Y)$ as:*

$$H(X|Y) = -\sum_{i}^{n}\sum_{j}^{m} P(x_i, y_j) \log_2 P(y_j|x_i) \tag{2.8}$$

The conditional entropy is closely related to the mutual information between the two random variables X and Y, that is the degree of dependence of the two distributions. Mutual information, usually written as $I(X,Y)$ is defined in relation to the entropy of the joint probability distribution of the two random variables. If the two random variables are independent, then $P(x_i, y_j) = P(x_i)P(y_j)$, however if they are not this identity is broken. One can define an entropy measure for each of the probability distributions $P(x_i), P(y_j)$ and $P(x_i, y_j)$, and the difference between them is defined as the mutual information. We write this as follows:

$$I(X,Y) = H(X) + H(Y) - H(X,Y), \text{ where} \tag{2.9}$$

$$H(X) = -\sum_{i}^{n} P(x_i) \log_2 P(x_i), \tag{2.10}$$

$$H(Y) = -\sum_{j}^{m} P(y_j) \log_2 P(y_j), \text{ and} \tag{2.11}$$

$$H(X,Y) = -\sum_{i}^{n}\sum_{j}^{n} P(x_i, y_j) \log_2 P(x_i, y_j) \tag{2.12}$$

It is easy to prove that $I(X,Y) = H(Y) - H(Y|X)$, and we can interpret the identity as saying that the mutual information between X and Y is the uncertainty in the value of Y *reduced* by the uncertainty in Y if we know the value of X. We now have the necessary elements to define Structural Entropy in terms of the interplay between the probability distribution of signal emission from a the vertex of the graph and the adjacencies of the graph as described earlier. Firstly, let us imagine a process whereby we randomly select a vertex from the graph, producing a probability distribution $P(V)$ for each vertex, which as the process of selection is uniform will be identically $\frac{1}{n}$ for each vertex in a graph of size n. Each vertex will in turn be a member of an independent set $s_i \in S$ (S is chosen to represent the independent sets to avoid confusion with I the mutual information). The conditional

probability $P(V|S)$ is the probability of selecting a given vertex, with knowledge of the stable set of which it is a member. These probabilities capture important information concerning the structure of the graph. Associated with $P(V|S)$ is a measure of entropy $H(V|S)$, or the uncertainty in the choice of vertex when the independent set is known. Using these quantities we define Structural Entropy as follows:

Definition 11. *The Structural Entropy of a Graph $G(V,E)$, over a probability distribution $P(V)$, $H(G,P)$, is defined as:*

$$H(G,P) = H(P) - H(V|S)\,, \tag{2.13}$$

where S is the set of independent sets of G, or equivalently the set of Chromatic Classes.

Definition 11 can be interpreted as the normal Shannon entropy of P, less the conditional entropy of a given signal occurring that is not distinguishable from any prior signal (and therefore being members of an independent set). Equation (2.13) is maximized when $H(V|S)$ is minimized. From the definition of conditional entropy in Definition (10), a minima is achieved when the conditional probabilities $P(V|S)$ are either very close to 1, or 0, which is the converse of the argument for maximization which is achieved when the probabilities are uniform (this is a standard result from the work of Shannon in [62]). We can interpret $P(V|S)$ as the probability, given a stable set S of the graph, of a randomly selected vertex $v \in V$ being a member of the stable set. For a connected graph, this is a property of graphs with the fewest edges such as the star graph, cycles and paths. For example the star graph on n vertices S_n, has two stable sets, one containing the 'hub' node, and one containing all of the other nodes. For the first stable set containing the 'hub' node these conditional probabilities are 1 for the 'hub node' and zero for all others, and for the other set it is zero for the 'hub' node and $\frac{n-1}{n}$ for all others. The absence of edges, per the construction of the graph, means that most symbols are not distinguishable from each other [63], and therefore little information is conveyed by any one signal. In our papers we prove that the star graph S_n maximizes structural entropy and the perfect graph K_n minimizes it, in the special case of each of the probabilities in P being equal.

2.2.2 Chromatic Entropy

Chromatic entropy is defined in terms of the subsets of V where each vertex in V has the same color label. These subsets, as defined in Definition 7, are called *Chromatic Classes* C_i, with the constraint that $\bigcup_i C_i = V$. Chromatic entropy is then naturally defined as:

Definition 12. *The Chromatic Entropy of a graph of n vertices is defined as:*

$$I_c(G) = \min_{\{C_i\}} \left[-\sum_i \frac{|C_i|}{n} \log_2 \left(\frac{|C_i|}{n} \right) \right], \tag{2.14}$$

where the minimization is over all possible collections of chromatic classes, or colorings, of the graph C_i.

This chromatic entropy is in fact closely related to the second term in Equation (2.13), $H(V|S)$, and if we assume that the probability distribution P is uniform, (an important assumption when applied to event management as this amounts to stating that no event is more likely to occur than any other), we can relate the two entropies with the following identity.

$$H(G) = \log_2 n - I_c(G) \tag{2.15}$$

2.2.3 Alternative Formulations

There are many other formulations of graph entropy, defined either in terms of matrix representations or graph ensemble considerations. Notable in the field of network science are two alternative definitions, which although in the work presented are not considered, we summarize here.

- *Von Neumann Entropy*: This definition of entropy utilizes the eigenvalues of the normalized Laplacian matrix of a graph. (For a good review see [57]). The normalized Laplacian matrix $\mathscr{L}$ is defined as $\mathscr{L} = \mathbf{D}^{-1/2}\mathbf{L}\mathbf{D}^{-1/2}$, where $\mathbf{D}$ and $\mathbf{L}$ are the normal degree and Laplacian matrix of the graph. In practice it has components $\mathscr{L}_{ii} = 1$, if $k_i > 0$, $\mathscr{L}_{ij} = -\frac{1}{\sqrt{k_i k_j}}$, if v_i and v_j are adjacent in the graph and 0 elsewhere. The definition of Von Neumann entropy is made in direct analogy to the quantum Von Neumann entropy. Once the eigenvalue spectrum of $\mathscr{L}$ is known, the Von Neumann entropy is computed as the negative log sum over eigenvalues, where λ_i is the i^{th} eigenvalue of the normalized Laplacian: $-\sum_i \lambda_i \log_2 \lambda_i$.

- *Randomized Ensemble Entropy* : Introduced in a series of papers by Bianconi, this definition builds upon the analogies between graphs and statistical thermodynamics first outlined by Newman and Park in [56]. In particular it considers general distribution functions for the degrees of nodes in a collection, or ensemble, of graphs sharing some common constraint (number of nodes, variance in degree and so on). The model then naturally permits analysis using tools of Statistical Thermodynamics such as Partition Functions which in turn lead to an Entropy of the ensemble. For a comprehensive survey see [3].

2.3 Dynamic Graphs and Graph Evolution

The initial work on vertex entropy presented in Chapters 4 and 5, led to an interest in the structure and degree distribution of the real world networks studied. This is an area that has been richly studied, and in the paper presented in Chapter 6, we make a novel contribution to the understanding of the degree distributions of real world networks. The study of these real world networks, dubbed 'Network Science', builds upon two fundamental models of network evolution and structure. For background we summarize the main points here, but a comprehensive treatment can be found in [14] and [2].

2.3.1 Random Graphs

The first models of dynamic graph evolution were the random models first introduced and extensively studied by Paul Erdős, and referred to as Erdős-Rényi or ER models. These graphs are constructed by first fixing the number of nodes $n = |V|$ and the number of edges $e = |E|$, and then randomly selecting e edges from the $\frac{1}{2}n(n-1)$ possible edges. Essentially models are built by choosing varying different schemes for the probability of an edge existing between any two nodes, and it is possible to compute many global properties of a graph such as the average degree of a node $\langle k \rangle$, and the probability distribution of node degrees $P(k)$.

Unfortunately when comparing these results with real world networks the models do not fit well. In particular the degree distributions typically exhibit a binomial distribution, which in the case of very large networks approximates to a Poisson distribution, whereas it has been long established that many real world networks have, at least an approximate, degree distribution that exhibits a power law such that $P(k) \propto k^{-\gamma}$ (first established by Faloutsos in [29]). The precise form of the degree distribution can often be much more complex than a simple power law, and variations are considered in more detail in section 2.3.3.

2.3.2 The Preferential Attachment Model

To address these shortcomings, Barabási proposed the preferential attachment model in a series of papers in collaboration with Réka Albert, [5], [1] and [2]. The model is built using the following assumptions:

- *Growth:* Starting with m_0 nodes and e_0 edges, we add a new node at each unit time step. When this node is added to the network, it connects to $m \ll m_0$ other nodes. This process continues indefinitely, such that after t unit time steps, there are $m_0 + t$ nodes, and $e_0 + mt$ edges. Eventually the constants in these expressions can be dropped as insignificant compared to t.
- *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.

The set up of the model proposes the probability of a *randomly chosen node i*, capturing a connection to a new node, as solely dependent upon its degree k_i compared to all other nodes as:

$$\Pi_i = \frac{k_i}{\sum_j k_j} = \frac{k_i}{2mt} \,, \tag{2.16}$$

Analysis then proceeds using a mean-field approach to finally arrive at a degree distribution expression:

$$P(k) = \frac{2m^2 t}{m_0 + t} \times \frac{1}{k^3} \,. \tag{2.17}$$

The detail of this model is presented in the paper in Chapter 6, but the great triumph of the approach is the emergence of a power law distribution with $\gamma = 3$. This model has gained widespread acceptance, although it has many points of disagreement with real data, those disagreements being the central focus of the research that is presented in our final paper in Chapter 6. In our paper, we modify Equation (2.16), introducing a new form of the attachment probability Π_i^c, as part of a new 'constrained attachment' model of network growth, that acknowledges a network connectivity limit. This does indeed lead to a better fit with real network data, but perhaps even more intriguingly, allows vertex entropy and preferential attachment to be linked in a more fundamental way.

2.3.3 Extensions to Preferential Attachment

Despite the success of the preferential attachment model, as noted above there are some known limitations of the model (elegantly outlined in Willinger *et al* [84]), which have motivated much research to propose additional models of network evolution. In particular there are three challenges that have received much attention:

- **Non Power Law Degree Distributions:** It is well established [84, 28, 19, 7] that not all networks have power law degree distributions, certainly not for all values of degree k. Accounting for these deviations requires significant modifications to the form of preferential attachment in Equation (2.16) that I will briefly outline below.

- **Younger Nodes can Have Higher Degree:** A natural consequence of the preferential attachment is that older nodes will acquire more links than younger ones [7]. Clearly in real world networks this is not the case. In the network of WWW site links, Google though much younger than say Yahoo!, nevertheless has acquired more links. Many attempts have been made to explain this including the fitness model of Bianconi and the introduction of initial attractiveness in the work of Dorogovtsev [7, 11, 27]. I will describe some of these modifications below.

- **Additional Attachment Factors other Than Degree:** A number of authors have attempted to introduce other factors than node degree into Equation (2.16) to explain the appearance of both scale free networks and the more complex form of degree distributions that are encountered in real world networks. These include approaches that are motivated by competition and game theory approaches (see for example Holme *et al* [34], the competition model of D'Souza [28] and the similarity model of Papadopoulos *et al* in [55]). In addition, in the paper presented in Chapter 6, I propose my own 'constrained model' of network evolution that alters Equation (2.16) by introducing a limitation of a node's capacity to accept further links.

- **Absence of a Physical Model for Preferential Attachment:** Despite much attention, the fundamental origin of Equation (2.16) is still relatively unaddressed. In its most basic sense it is an observation of network evolution dynamics rather than an axiom of nature. This has motivated many to explore whether there are other more fundamental processes at play. Most often these draw inspiration from statistical mechanics, such as the exponential random graphs studied by Newman and Park in [56], or indeed generalizations of the fitness models of Bianconi in the Network Geometry with Flavor (NGF) models that explore the attachment dynamics at higher dimensions in [12, 21].

In the paper I present in Chapter 6, I put forward an argument based upon vertex entropy that seeks to derive the form of Equation (2.16).

Alternate forms of the Preferential Attachment Probability

The work of Krapivsky *et al* [42] introduces a general exponent to the degree term in Equation (2.16), modifying it to k^{α}, where $0 < \alpha < \infty$. This introduces several regimes to the degree distribution the 'sub-linear' and 'super-linear' regimes. For $\alpha = 1$, the standard preferential attachment model is recovered. In the 'super-linear' region the network collapses to a hub and spoke model with one node capturing a connection to every other node. In the 'sub-linear' region $0 < \alpha < 1$, the degree distribution is a power law with an exponential cut-off. This model is effectively tunable to produce the degree distribution observed, but is not capable of explaining why a young node can be more attractive than an older one. An early attempt to extend the model to include 'initial attractiveness' was made by Dorogovtsev in [27], in which a node is assigned a constant initial attractiveness when it is added to the network. The model produces a more complex degree distribution than the standard preferential attachment model, but requires an arbitrary attractiveness parameter that is constant across all nodes. This model was extensively extended in the fitness model of Bianconi, described below.

In the work of Zhou *et al* [88], a more complex form of the non linear attachment probability is introduced, motivated by consideration of positive feedback in the evolution of the internet. The model uses a modification of the random graph ER model, in which internal links are introduced between existing nodes, with a probability less than the probability that the new node connects to a given existing node. This model is parameterized by a form of Equation (2.16), where node degree k_i has an exponent $1 + \delta \log_{10} k_i$, where δ is a free parameter. In the Internet Autonomous System network, empirically the exponent has a best fit when for all nodes $1 + \delta \log_{10} k_i = 1.166$. This is an interesting result, reproducing some of the same degree distribution behavior seen in the work of Krapivsky.

The model of Zhou also introduces a hybrid between preferential attachment and random model evolution, which is in some way similar to the empirical model of Tian Bu *et al* [76], and indeed the link copying models of Donato *et al* [26]. In all of these models, in addition to a probability of attachment arising from a node's degree additional links are added or removed subject to a random probability. The justification for the addition and removal of links arises from considerations of tendency for new nodes in a network such as a web graph, to replicate the local topology of nodes that they link to, or for links to become deleted as web sites are decommissioned.

Models Inspired by Competition and Physical Models

A number of approaches to extend Equation (2.16) have been taken that draw their inspiration from different models of node attractiveness. These include a broad range of techniques such as optimization theory in the work of D'Souza *et al* [28], that seek to explain network evolution as the solution to an optimization problem. The optimization occurs on a cost function that seeks to balance closeness of any two given nodes with the cost of connecting a distant node to the centre of the network. This model produces an exponentially corrected form of degree distribution with $P(k) \propto k^{-\gamma} e^{-\alpha k}$. In addition, the work of Holme *et al* [34] attempts to model network growth as an optimization in the traffic carrying capacity of the network subject to the economic constraints of adding additional network infrastructure. Using numerical simulations it is able to accurately reproduce the degree distributions and spatial distribution of the internet. Perhaps the most widely known model that draws from considerations of competition is the 'fitness' model of Bianconi described in [7, 11]. This model parametrizes the attractiveness of the node using a *fitness* measure, η_i, which is fixed, or quenched, at the time of introduction of the node into the network and drawn from a supplied probability distribution $\rho(\eta)$. The form of Equation (2.16) is modified to:

$$\Pi_i = \frac{\eta_i k_i}{\sum_j \eta_j k_j} \,. \tag{2.18}$$

In their analysis this form of attachment can be solved for a class of distributions $\rho(\eta)$ analytically, and successfully reproduces scale free networks with power law exponents $2 < \gamma < 3$ and an exponential cut-off. The model produces a degree distribution $P(k) \sim \frac{k^{-1+C}}{\log(k)}$, where C is a constant . In this model a deep analogy with statistical mechanics can be made by re-defining the fitness parameter as $\varepsilon_i = -\frac{1}{\beta}\log \eta_i$, with β being identified as classical inverse thermodynamic temperature, and the 'energy level' ε_i is populated by a 'particle' whenever there is a connection to node i. The denominator of Equation (2.18) is then easily identified with the partition function Z, familiar from the Bose-Einstein model of statistical mechanics. Further, it is demonstrated that networks can exhibit behavior similar to Bose-Einstein condensation where all particles collapse into a single energy level. This behavior is exhibited when a single node in the network acquires connections to all other nodes and is an example of the 'super-linear' behavior of the Krapivsky model when $\alpha > 2$.

Recently the model has been significantly extended by Bianconi *et al* [12, 21], to multi-dimensional forms of attraction where the interaction is not between nodes but between groups or simplexes of nodes. A simplex is a fully connected clique of nodes, the number of

nodes being equal to $d+1$ where d is the 'dimension' of the model. This has deepened the analogy with both statistical mechanics and some forms of quantum gravity. It is an exciting development of the network science journey and points to further richness in the preferential attachment model.

Models with Additional Factors to Node Degree

The fundamental premise of preferential attachment is that node degree is the sole contributor to a node's ability to attract new connections. In the work of Papadopoulos *et al*, it is argued that similarity could be another factor in a node's attractiveness. Similarity is introduced as a free parameter that measures a node's compatibility to a new node in the network in terms of interests. For example, when a new blog site is created, it is likely to be linked to blog sites that the author knows as well as popular websites, such as search engines, with many connections. The model developed from this basis is again capable of producing scale free degree distributions with power law exponent in a range $2 < \gamma < 3$.

In the paper presented in Chapter 6, the approach taken to produce a model with more realistic degree distributions introduces the concept of a node capacity. This capacity will be familiar from the design of real networks, where a network switch, for example, has a finite number of connections it can support. The model produces a scale free distribution of node degree, at least for small values of k, but the evolution of a node's degree is capped by the capacity constraint.

Models with additional constraints are capable of producing better fits to the degree distributions of real world networks, and there are many possible potential modifications. They still do not, however, explain the origin of node attractiveness, an important open problem.

Chapter 3

Overview of Published Work

3.1 Towards and Approximate Graph Entropy Measure for Identifying Incidents in Network Event Data

This paper was prepared for the IEEE NOMS 2016 AnNet workshop, and was accepted for presentation and publication. It contains the first set of results that were obtained using the vertex entropy technique to analyze the topology, event and incident data harvested from a customer of my employer Moogsoft Inc, and from the Internet Topology Zoo [40].

The initial definitions of vertex entropy were somewhat modified in later publications, but contain the essential approach of combining locally definable properties of a node to create measures of vertex entropy that can be used to filter out events from unimportant nodes. I extended the definitions for node level entropy in [23] and [24] to obtain the form for the vertex entropies used in the paper and the initial results from the real world data gave strong indication of a correlation between high values of vertex entropy and a higher tendency for a node to produce incidents.

The paper (described in Chapter 4) was presented at the conference in Istanbul and was awarded 'Best Paper' for the workshop.

3.2 Vertex Entropy as a Critical Node Measure in Network Monitoring

Following on from the conference paper, I significantly extended the analysis of the datasets, and outlined the theoretical treatment of the measures in much more detail. This included

proving that the vertex entropy definitions act, at least in their limiting behavior, as good analogs of the global entropy variants. I was also able to prove that as an entropy measure they obey the requirements of *additivity*, *maximality*, *positivity*, and *symmetry*.

Initially this was submitted as a main track conference paper for INFOCOM 2017, but was rejected. The comments received were used to shape the full paper presented in Chapter 5. In particular I extended the entropy measures to investigate the use of clustering coefficient as a node probability and also introduced F_β measures, based upon F_1 scoring (described in [59]), to strengthen the experimental analysis.

The paper after two rounds of review has been accepted for publication in the IEEE Transactions on Network and Service Management and is currently in the production process. It is available by following the pre-publication reference here [74].

3.3 Constraints and Entropy in a Model of Network Evolution

During the research for the first two papers, I became interested in the degree distributions of real world networks. This was initially motivated by discussions I had at the Mathematics of Networking 15 conference at Oriel College. In particular I noticed that the commercial network data did not follow a scale free distribution as would be expected from [2], and other standard works.

For communications networks, the presence of physical and logical constraints in their construction is at odds with the fundamental assumption in the standard treatments of preferential attachment that a network node can indefinitely acquire connections. This led to investigation of an extension to the theory of scale free networks, which I call *Constrained Attachment*, using a simple uniform constraint on the maximum degree of a node. As this yielded good early indications of a better fit to the data, the model was properly developed and presented in the submitted paper.

As an addendum, I also became interested in ways in which the dynamical evolution of networks could be related back to concepts of entropy, introduced in the earlier publications. The paper concludes with an intriguing result that shows how vertex entropy could be used to build a scale free model of network evolution.

These results were collected together into the paper presented in Chapter 6, and is currently under review by the editors of European Physical Journal B. Initial submission to Physical Review E was rejected after two rounds of review, although one of the reviewers

recommended publication. After consultation with my co-authors we felt that the manuscript should be submitted to another high impact journal and European Physical Journal B was chosen.

3.4 Other Submissions and Conference Talks

During the course of my research I gave three research talks, presenting the results of my work. The first of these was at the IEEE NOMS 2016 conference in Istanbul, where I presented the paper described in Chapter 4.

This talk was extended into a work in progress seminar I delivered at Sussex in June of 2016, including early results from the work on constrained attachment. Finally I applied to, and was accepted, as a speaker at the Mathematics of Networking 15 conference at the University of Bath. At that conference I presented the current state of the constrained attachment model, and met Jonathan Dawes, which began a collaboration which culminated in his participation in the paper presented in Chapter 6.

Chapter 4

Towards and Approximate Graph Entropy Measure for Identifying Incidents in Network Event Data

4.1 Background to First Publication

4.1.1 Motivation and Summary of Contribution

The original motivation for the research was to identify novel, easy to compute measures, of node importance in a communications network that could be used to eliminate noisy events. This is a significant problem when deploying fault localization software at scale.

The original work was conducted upon toy graph models such as in Figure 2.1, with attempts to try and single out nodes that have critical importance to the connectivity of a network. This work resulted in the first forms of vertex entropy and a poster paper was prepared for SIGCOM 2015. This poster was rejected, mostly due to the work not having any justification from application to real data.

The work to produce this analysis was underway but not mature enough to present in the poster. The paper submitted to the AnNet workshop at IEEE NOMS 2016 was the first opportunity to collate and present this work, together with the results obtained from applying the methods to the data we had obtained from a large web portal customer of my employer. For confidentiality reasons I am not able to identify the company and the data can only be used for analysis and is not publicly shareable. We are working to anonymize the data sufficiently, but at this time it is still only accessible under a strict confidentiality undertaking.

Particularly pleasing was that not only was this paper selected for the conference, but it went on to win 'Best Paper' at the end of the session.

4.1.2 Theoretical Contribution

The starting point for the analysis was the concept of node importance. There are many available measures of importance, mostly based around the concept of centrality ([65], [82], and [50] are good overviews). One particular measure of complexity in graph theory had received little attention, namely entropy.

The history of entropy as a measure of information in computer science dates all the way back to Claude Shannon's breakthrough work (see [62] for details), where the concept of entropy is related back to the probability of certain signals being emitted, and measures the so-called relative *degree of surprise* when receiving a given signal. In a similar way a graph can be considered to contain information about how nodes are connected to each other. The basic supposition is that a node that is connected in an unusual (i.e. a high entropy) way could be more important as it would indicate that the node is in some way carrying more information about how the network is constructed.

The significant impediment to this is that all of the measures of entropy for graphs that are well understood are global, have nothing to say about an individual node, and are expensive to compute (this is discussed in this and the following publication, but the relevant references are [41], [63], [57] and [50]).

In this paper, we build upon the work of Dehmer [23, 24] to introduce valid local measures of vertex entropy. We do not, in this publication, explore the detail of the theoretical treatment due to constraints on the length of the paper, but this is covered in much more detail in the publication described in Chapter 5. What we are able to demonstrate is that against the event, incident and topology data we have from a real world communications network, these measures are successful in identifying nodes that are more likely to surface incidents, which is the conclusion of the paper.

4.1.3 Data and Methods Used

The data that was analyzed came from two principal sources. The first was an academic repository of the network topologies of large service providers, the 'Internet Topology Zoo' (ITZ) [40], and the second a large operational dataset from the web portal customer. The ITZ is a useful repository of varied real world networks against which to test the operation of the analysis harness, and also to investigate how effective the algorithms were at identifying

critical nodes in an easy to visualize way. The ITZ comes with a suite of online visualization tools, from which we extracted the images in Figures 9 and 10 of the paper. Unfortunately this dataset is static and contains no operational data relating to faults and incidents.

The commercial data contains a rich source of event and incident data, and in particular allows the analysis of event and incident rates in the context of their originating nodes. The principal goal of the research was to obtain data relating to the distribution of event and incident data by entropy metric of the originating node. To achieve this a range of software tools were written to process the source data. They were implemented in JAVA, and operated in conjunction with a MySQL database for permanent storage. The analysis programs built included:

- `graph_analyser`: This executable was built to ingest source topology as a list of edges in either a Comma Separated Values (CSV) or GraphML (details on this format are available here [33]). The executable could also optionally load data from a table containing a list of edges. The program could be directed to produce tables of all of the entropy metrics, and also distribution tables of degree, clustering coefficient, and each of the defined entropies. These tables were stored in the database for later query, both for analysis and use in the other programs.

- `event_analyser`: This executable can ingest an individual event file, or read a directory for all suitable event files. Each file contains a list of events with attributes separated by a '|' symbol, as supplied by the customer. The analyzer reads these files and maps each field to a token, which is subsequently used to create records in the database for each event. The event format follows the general pattern as described in Chapter 1, and contains a record of the originating node along with other useful descriptors of the event condition. The analyzer performs basic duplicate suppression using a hash key built from a subset of fields in the event to mirror the operation of a typical fault management system (see [36]). As part of the record stored in the database the analyzer looks up the values of the various entropy metrics, clustering coefficient and degree, and also builds an event count distribution for each of them.

- `incident_analyzer`: This executable operates in an almost identical fashion to the event analyzer, but instead analyzes data that is obtained from a report ran on the customer's incident management system. This system records every escalation of an event to successive layers of technical support and records for each incident the level of escalation. The incident analysis harness extracts a subset of these escalations, according to the customer's process, to include those incidents that were deemed

sufficiently impacting to warrant escalation beyond the first layer of triage. In a similar way to the event analyzer the output is a record of distribution of incidents by the various entropy measures, clustering coefficient and degree. These values are obtained from the node associated to the incident, being used to look up the values from the output of the `graph_analyser`.

Once the analysis programs were run, results were extracted and plotted to produce the results presented in the paper.

4.1.4 Contributions from Co-Authors

Although the work outlined in the paper is substantially my own, both of my co-authors contributed much in helping to shape the argument presented, checking and critiquing the manuscript and shepherding me through the submission process.

4.1.5 Related Work

In a series of articles from 2004 onwards, Jon Stearley *et al* of Sandia Laboratories published details of the Sisyphus log management tools([66],[67], and [54]). In these papers Stearley describes a system that uses term frequency to construct an entropy measure for each entry in a system log, and then uses that measure to prioritize the log message. A derivative of this approach has been successfully developed at my company Moogsoft Inc, and is the subject of patents I authored. It is used to assist in the automatic elimination of noise in event feeds.

This method, using characteristics intrinsic to the event data to suppress the bulk of events that are not causal, was the starting point for the program or research. The key difference that vertex entropy has to event entropy is the use of auxiliary topology data to provide context to the significance of events. Rather than focus upon the information carried in the text of the events, I chose to focus on the information contained in the structure of the network. This approach I believe is completely novel, and to date there is no published work that directly implicates the graph entropy of a network in the operational significance of events.

Towards an Approximate Graph Entropy Measure for Identifying Incidents in Network Event Data

Phil Tee
Moogsoft Inc
140 Geary Street, San Francisco, CA 94108
phil@moogsoft.com

George Parisis and Ian Wakeman
School of Engineering and Informatics
University of Sussex
Brighton, UK
{g.parisis, ianw}@sussex.ac.uk

***Abstract*—A key objective of monitoring networks is to identify potential service threatening outages from events within the network before service is interrupted. Identifying causal events, Root Cause Analysis (RCA), is an active area of research, but current approaches are vulnerable to scaling issues with high event rates. Elimination of noisy events that are not causal is key to ensuring the scalability of RCA. In this paper, we introduce vertex-level measures inspired by Graph Entropy and propose their suitability as a categorization metric to identify nodes that are *a priori* of more interest as a source of events.**

We consider a class of measures based on Structural, Chromatic and Von Neumann Entropy. These measures require NP-Hard calculations over the whole graph, an approach which obviously does not scale for large dynamic graphs that characterise modern networks. In this work we identify and justify a local measure of vertex graph entropy, which behaves in a similar fashion to global measures of entropy when summed across the whole graph. We show that such measures are correlated with nodes that generate incidents across a network from a real data set.

I. Introduction

An important objective when monitoring a large scale network is detecting failures in critical nodes. This is accomplished by collecting notifications or *events* from the network and analysing these events to determine failed nodes. Events occur at a high rate, and do not always directly indicate a problem. To illustrate, at a typical large enterprise network[1], the event rate is 135 million events a day, generated by just a few hundred 'actionable incidents'.

Identifying which events are the cause of actual outages is called Root Cause Analysis (RCA) [1]. Many algorithms are used to perform RCA [1], but scalability limitations make applying these algorithms to the full event stream impractical. To perform RCA across all events, the flow of events has to be significantly reduced (for example see [2]).

The most common approach to reducing the event rate is the simple act of discarding uninteresting events with a manual filter or exclusion list, a process known as 'blacklisting'. This process is extremely time consuming and error prone. At industrial scale, blacklisting can require thousands of rules; in a fast changing network, such an approach is not practical. A method to automatically eliminate uninteresting events would yield significant savings.

In this paper, we introduce a novel technique derived from Graph and Information Theory that determines, which events can be treated as noisy based on the location of their source in the network. The technique relies upon the use of Information Entropy [3], and Graph Entropy [4], [5]. We hypothesise that nodes contributing most to the entropy of a graph are the nodes most likely to generate incidents when events occur. An alternative formulation of the problem is that those nodes contributing most to the connectivity of a graph are most likely to generate incidents when they fail. Graph Entropy is, however, computationally expensive, so we propose alternative formulations that provide similar properties to graph entropy but can be calculated using known global graph properties and information local to the node. We demonstrate that these measures correlate well to the node event pairs that result in incidents.

II. Network structure and other work

Following the influential analysis of Barabàsi and Albert [12], there was much work investigating the structure of communication networks, such as by Faloutsos et al [7] and Li et al [8]. The approach primarily focused on datasets generated by discovery protocols such as `traceroute`. This approach was used by Barabási and Albert to assert that communications networks have a power law node degree sequence, possessing the *Scale-Free* property, whereby, node degree distributions obey the inverse power distribution law. This was further used to justify the claim that communications networks, like the Internet, are both robust to random attack and vulnerable to targeted attack (the central arguments are outlined in [9], [10], [11], and again in [12]).

The drawbacks of `traceroute` as a discovery protocol are well understood, and outlined clearly in [13] and [14], but essentially arise from the fact that the nature of the `traceroute` tool hides network structure at protocol layers other than IP, and creates many false, high degree nodes. Using more accurate data, built manually from operational change tracking databases of real world networks, is a far better way to analyze networks for vulnerability, and includes true connectivity not confined to the IP protocol. We have gained access to a number of datasets from customers of Moogsoft,

[1]This work is underpinned by the experience at Moogsoft in supplying large scale network management software to many blue chip customers.

978-1-5090-0223-8/16/$31.00

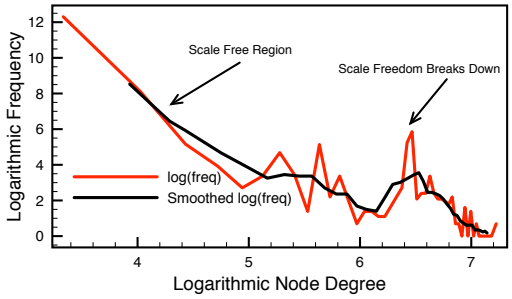


Fig. 1. Scale Freedom Breakdown in a Real Network of 225,239 nodes.

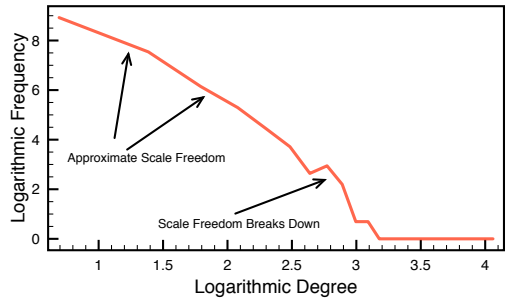


Fig. 2. Scale Freedom Breakdown In the Internet Topology Zoo.

which number in excess of 200,000 devices and cover many autonomous networks. We can easily dispel the notion of simple power law degree sequences, and hence the generalities implied in [12] and [11], with this dataset, as illustrated in Figure 1 and using the network data published in the Internet Topology Zoo [14] in Figure 2. What is evident from the degree distribution analysis is that at best the power law is an approximation at low degree, with significant deviations as degree increases. Furthermore for the proprietary dataset the distribution has a notable cluster at high degree values.

Nevertheless, this approach of analyzing communications network using graph invariants such as node degree and other related metrics, does indicate that there are methods of identifying nodes, which are of more interest from a network vulnerability perspective. The individual contribution of a network node to the overall connectedness of a network, and hence the potential impact of that node failing is an important problem in network management, and the subject of much commercial activity. This has typically been confined to behavioral models of the network (see for example [1], [15]), but these are susceptible to poor scaling behavior on large networks where changes in network topology are frequent. This has particular impact in current networking technologies such as SDN (Software Defined Networking), a compelling illustration being [16].

Much focus has been spent in the literature on degree based characterizations of networks from an analysis basis, but it is accepted that degree sequences do not uniquely determine the connectivity properties of a network. Indeed the determination of metrics that allow two networks to be compared for similarity is a much studied and challenging problem in graph theory ([6], [17]). It is the object of this work to establish whether there are other, deeper, node level metrics that can identify the important nodes in a network and yield a useful operational tool to identify operational vulnerabilities of communications networks.

III. Towards Local Measures for Graph Entropy

Historically, entropy has been defined in Graph Theory[2] as a measure of complexity of the global structure of a graph. As a metric it captures many important characteristics, which are of direct interest in a number of applied fields, including the analysis of failure modes of communication networks. In particular, networks with non uniform connectivity will have high values of entropy. Unfortunately the three most well understood measures of entropy involve calculations that have impractical computational complexity, as a graph scales in terms of the number of vertices and edges. What is worse, any change to either the edges or vertices of a graph requires an entirely new computation across the whole graph, and it is extremely difficult to compute the contribution of an individual node to the entropy of the graph. The three variants of Graph Entropy that we shall concern ourselves with are:

- **Chromatic Entropy**: Chromatic entropy is defined by partitioning a graph into sets (or colorings) of disconnected vertices.
- **Körner or Structural Entropy**: The original entropy measure defined on a graph, intended to capture the mutual informational of stable sets.
- **Von Neumann Entropy**: Introduced in analogy to the entropy of quantum systems, this is defined against the eigenvalues of the *Laplacian* matrix associated to a graph.

A valid entropy measure is expected to satisfy the following conditions: *maximality*, *additivity*, *symmetry* and *positivity* [4], [18].

In our treatment we make reference to a number of special graphs, which we define here as:

- K_n **The Complete Graph**: This graph is formed from a set of n vertices, maximally connected.
- S_n **The Star Graph on** n **Vertices**: This graph has one vertex v which is connected to all other vertices, with no other edges in the graph.
- P_n **The Path on** n **Vertices**: This graph is a simple chain of n vertices with no loops, and a start node v_1 and an end node v_n.
- C_n **The Cycle on** n **Vertices**: This graph is a special case of P_n such that $v_1 = v_n$; each node has degree 2.

A central objective of our work is to establish easily computable metrics that measure the contribution of an individual node to the entropy of the whole graph. We will establish that the values, when summed across the whole graph give values consistent with the global measures, and have minimum and maximum values for the same types of special graphs.

[2]We follow standard graph theory notation for edges and vertices in our presentation.

This analysis establishes the proposed metrics as candidates for local vertex entropy measures, and in further work we investigate the relationship between the metrics in more detail.

Recent work by Dehmer on Graph Entropy [19],[20] provides a framework that unifies the three global invariants discussed, and provides a pathway to extend these measures in a more computable direction. In particular, both Structural and Chromatic entropy rely upon partitions of the vertex set of the graph, which are known *NP-Hard* problems, and, Von Neumann Entropy requires an expensive computation of eigenvalues for the Laplacian Matrix of the graph.

Dehmer defines the concept of a local functional for a vertex, which can be scoped to calculate values for every vertex based upon the local topology of the graph. The degree of locality in the treatment is controlled by using the concept of j-spheres, S_j in the graph, centered at a given vertex. Formally if we denote by $d(v_i, v_j)$ the shortest distance between nodes v_i and v_j, the definition of a j-sphere proceeds as follows:

Definition 1: For a node $v_i \in V$, we define the 'j-Sphere' centered on v_i as:

$$S_j = \{v \in V | d(v_i, v) = j, j \geq 1\} \tag{1}$$

and for convenience later, the related 'j-Edges' E_j as

$$E_j = \{e_{ij} \in E | v_i \in S_j \text{ and } v_j \in S_j\} \tag{2}$$

Using this definition, we then equip each S_j with a positive real-valued function $f_i : S_j \to \mathbb{R}^+$, and further construct a probability functional for each vertex as

$$p_i = \frac{f_i}{\sum_{v_j \in V} f_j} \tag{3}$$

which trivially satisfies $\sum_i p_i = 1$.

The principal direction of Dehmer's proposition is that these functions f_i can be constructed from any structural measure valid and calculable within the 'j-Sphere' of a vertex. In the published work [19],[20], however, these functions are somewhat complex expressions, which introduce global invariants of the graph complicating their computation. We now move on to the important result of this paper, which is the introduction of a variant of Dehmer's approach that uses purely local properties of the neighborhood subgraph of a vertex, and global constants of a graph, such as, the number of nodes n or the number of edges $|E|$.

A. Local Vertex Entropy Measures

Given the constraint of locality, a number of constructs can be designed that satisfy the probability functional defined in equation (3) up to a normalization constant. It is possible to define the notion of locality using the general concept of j-spheres, but in our treatment we restrict the constructions to a 1-sphere for simplicity of explanation. In the immediate neighborhood of a vertex the available measures are restricted to the degree of the vertex and the presence of cycles in the local subgraph. It is important that the measures that are constructed are bounded in an acceptable way, when summed

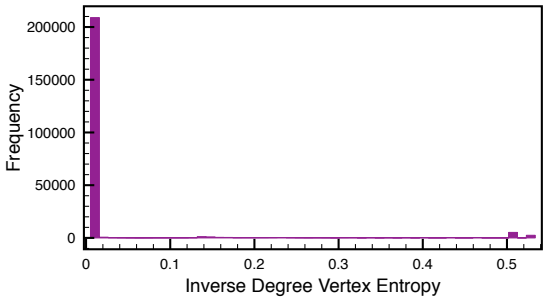


Fig. 3. Frequency Distribution of $VE(v)$.

across the whole graph and satisfy the fundamental properties of an entropy measure: *maximality*, *additivity*, *symmetry* and *positivity* [4], [18].

1) Inverse Degree: The first, and most basic probability functional that we can construct on the 1-sphere of a vertex is its inverse degree. In this case we write the probability at a vertex as:

$$p_i(v_i) = \frac{1}{d_{v_i}} \tag{4}$$

and the corresponding entropy of the vertex $VE(v_i)$, and whole graph $H_{InvDegree}$ as

$$VE(v_i) = \frac{1}{d_{v_i}} \log_2(d_{v_i}),\ H_{InvDegree} = \sum_{i=0}^{i<n} \frac{1}{d_{v_i}} \log_2(d_{v_i}) \tag{5}$$

The first observation is that the sum of inverse degrees does not satisfy the constraint $\sum_i p_i = 1$. However, one can observe that for any given graph G, this probability functional sums to the constant:

$$C = \sum_{i=0}^{i<n} p_i = \frac{\sum_{i=0}^{i<n} \left(\prod_{j \neq i} d_j \right)}{\prod_{i=0}^{i<n} d_i} \tag{6}$$

We note that $p_i = C \times \frac{1}{d_{v_i}}$, and discard the constant as part of the normalization.

We can, however, establish bounds for the value of $H_{InvDegree}$, algebraically. As $p_i < 1$, we can expand (5) to obtain:

$$H_{InvDegree} \approx -\sum_{i=0}^{i<n} \frac{1}{d_{v_i}} \left(1 - \frac{1}{d_{v_i}} \dots \right) \tag{7}$$

Firstly the value is maximized in the case of all degrees being equal and at their maximum. This is satisfied by the complete graph K_n. The minimum requires that the average degree for the graph is at a minimum. This is satisfied by the star graph on n vertices, S_n.

Using the same collection of experimental data used to generate Figure 1, we plot the distribution of values of Inverse Degree Entropy for all nodes in Figure 3. The presence of a large number of edge nodes of degree 1, heavily skews the

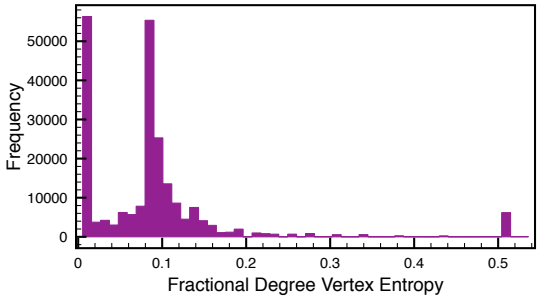


Fig. 4. Frequency Distribution of $VE'(v)$.

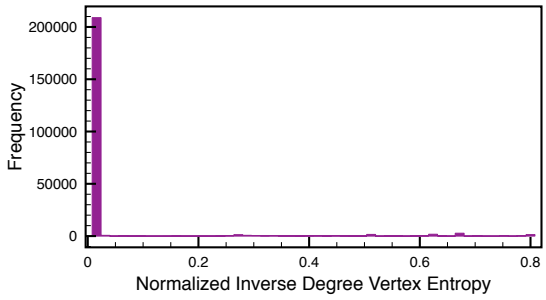


Fig. 5. Frequency Distribution of $NVE(v)$.

distribution, but there is a significant cluster of nodes at a value of 0.5.

2) Fractional Degree Entropy: Inverse degree is unsatisfactory in some regards. Firstly the probability functional is not naturally defined to satisfy the unity sum constraint. Secondly, and more importantly, the degree of a vertex does not capture how '*hub-like*' the node is relative to others. To capture this, we can define an alternative functional, which is based upon the ratio of the vertex degree to the total number of edges in the graph, as follows:

$$p_i(v_i) = \frac{d_{v_i}}{2|E|} \tag{8}$$

Given that $\sum_{v \in V} d(v) = 2|E|$ this functional directly satisfies the unity sum constraint. In a parallel way to equation (5), we define the fractional degree entropy as:

$$VE'(v_i) = \frac{d_{v_i}}{2|E|} \log_2 \left(\frac{2|E|}{d_{v_i}} \right) \tag{9}$$

$$H_{FractDegree} = \sum_{i=0}^{i<n} \frac{d_{v_i}}{2|E|} \log_2 \left(\frac{2|E|}{d_{v_i}} \right) \tag{10}$$

Following the treatment of Inverse Degree Entropy we establish bounds on this measure by considering the extremal graphs K_n and S_n,P_n. If we expand the logarithmic term in equation (9) we obtain the following higher order term for $H_{FractDegree}$:

$$H_{FractDegree} \approx \sum_{i=0}^{i<n} \left\{ \frac{d_i^2}{|E|^2} - \frac{d_i}{|E|} \right\} \tag{11}$$

This is minimized for the graph over n vertices with minimum degree sum P_n and maximized by K_n.

We plot this value distribution in Figure 4. The distribution of the values is more spread out compared to the Inverse Degree Entropy, but still shows the '*Double Bump*' feature with a cluster centered around a value of 0.1, and a smaller cluster around 0.5. The presence of this '*Double Bump*' in both measures is a necessary but not sufficient condition for these metrics to be useful in highlighting nodes whose impact on connectivity is proportionately higher than others, as both cleanly segregate the nodes into two sets of high and low vertex entropy.

3) Normalized Degree Entropy: There is a considerable practical difference between a star network topology and a fully meshed one, that is, between S_n and K_n. In the former, the network is vulnerable to the loss of its central high degree vertex; in the latter, the loss of any one vertex can never create isolated vertices. Both prior measures make no distinction between these two topologies, but there are available metrics measurable at one hop distance that capture this concept. Introduced in [21] and [12] is the concept of the clustering coefficient of a vertex. Utilizing the degree of the vertex i, d_i, it is possible to calculate the fraction of possible edges in the local neighborhood and thereby define the clustering coefficient as:

$$C_i = \frac{2|E_i|}{d_i(d_i+1)} \tag{12}$$

This metric captures how well meshed a node is into its local neighborhood, and therefore serves as an ideal candidate for further refining the vertex measures introduced earlier. In essence, we want to highlight vertices whose clustering coefficient is low, that is, their local neighborhood is more similar to S_n locally than K_n. To that end we define the following *Normalized* Vertex Entropies:

Definition 2: We define for a graph $G(V, E)$ the following *Normalized* Inverse Degree Entropy for both vertex and total graph as follows:

$$NVE(v_i) = \frac{1}{C_i} \times VE(v_i) \tag{13}$$

$$H_{NormInvDegree} = \sum_{i=0}^{i<n} \frac{(d_{v_i}+1)}{2|E_i|} \log_2(d_{v_i}) \tag{14}$$

and the corresponding definition for fractional vertex entropy is defined similarly:

$$NVE'(v_i) = \frac{1}{C_i} \times VE'(v_i) \tag{15}$$

$$H_{NormFractDegree} = \sum_{i=0}^{i<n} \frac{d_{v_i}^2(d_{v_i}+1)}{4|E||E_i|} \log_2 \left(\frac{2|E|}{d_{v_i}} \right) \tag{16}$$

Using similar arguments to those used for the non-normalized versions, it is simple to verify that these quantities

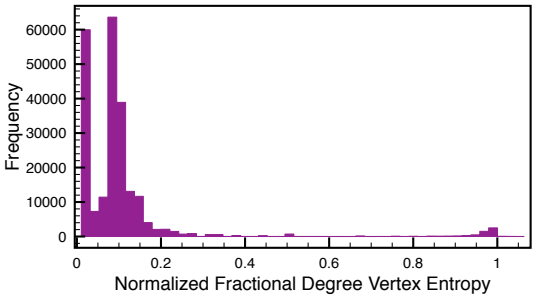


Fig. 6. Frequency Distribution of $NVE'(v)$.

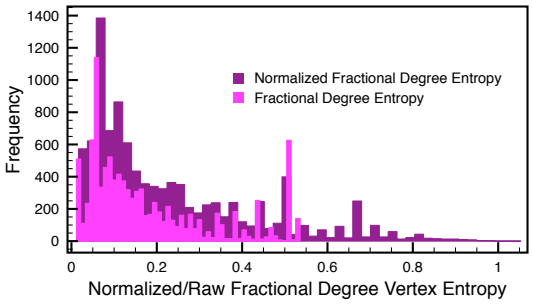


Fig. 7. Frequency Distribution of $NVE'(v)$ & $VE'(v)$ for the Internet Topology Zoo.

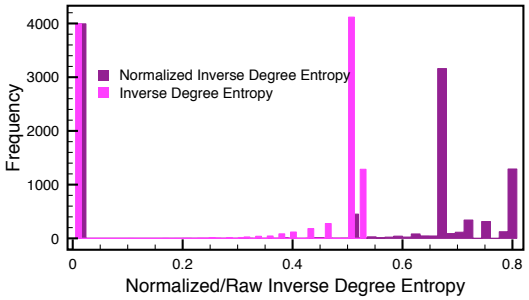


Fig. 8. Frequency Distribution of $NVE(v)$ & $VE(v)$ for the Internet Topology Zoo.

TABLE I
VALUES OF NORMALIZED ENTROPY FOR SPECIAL GRAPHS

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

are minimized by the graph P_n, and, maximized by the complete graph K_n.

For the same dataset used previously, we plot the distributions of these quantities in Figure 5 and Figure 6. It is interesting to note that both quantities share the same '*Double Bump*' distribution as the non-normalized forms, with a more pronounced separation of the two clusters. We can apply the same analysis to the data in the Internet Topology Zoo [14] and we obtain the distributions in Figure 7 and Figure 8. Although the Internet Topology Zoo is a smaller dataset (19,476 vertices in total) than the proprietary dataset, it still demonstrates a noticeable cluster at high values of both the normalized and raw values of vertex entropy. This '*Double Bump*' style distribution is a necessary, though not sufficient, feature of this metric for it to be useful as a tool in identifying nodes of crucial importance in network monitoring.

To illustrate the bounding values of these normalized quantities for our normalized entropies, summed across our special graphs, we summarize the values in Table I.

From this it is possible to conclude that for NVE, C_n maximizes the value, whereas, S_n minimizes it, and for NVE' P_n gives the maximum value and K_n the minimum.

IV. CONCLUSIONS

The main aim of this paper is to introduce computable, node level alternatives to structural entropy measures that are defined across the whole graph. Inspired by the advances made in Barabási's pivotal paper, and suggestions made in the work of Dehmer, we have advanced two computable metrics using structural information available within one hop of a network node. Both of these measures we applied to the proprietary data set, and, to the Internet Topology Zoo data, in both a raw and normalized form. The normalization adjusts the degree based values by the extent to which the local neighborhood of the node is clustered. When these values are applied to the datasets we obtain a distribution, which contains two peaks in value, the second peak at higher values of the metric involving far fewer nodes than the first.

The utility of these local measures of entropy *requires* such a distribution if it is to be effective at identifying specific nodes in the networks, which introduce vulnerability to the network in terms of overall connectivity. This is more precise than simply selecting the nodes of highest degree, which is central to the scale free argument that a few highly connected nodes, well chosen, represent the bulk of the vulnerability of a network. Nodes with high degree may be critical to the functioning of the network, but are equally likely to be in a highly meshed and therefore redundant part of the topology. It is the purpose of the normalization of the vertex entropy values to suppress high degree nodes in highly meshed parts of the network over high degree nodes, which are less redundantly wired into the network.

The ultimate test of these values is to examine failure modes in real networks, and, identify if a high value of $NVE(v)$ or $NVE'(v)$ does correlate with those nodes whose failure, and removal, cause more operational impact on the functioning of the network. For that purpose, we have analyzed the commercial datasets we have access to at Moogsoft and present in Figure 11 an encouraging indication of the utility of one of our measures $NVE'(v)$. We analyzed the distribution

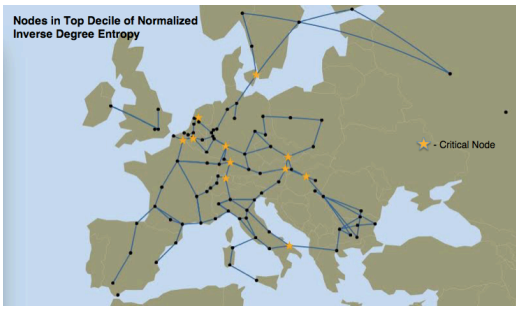


Fig. 9. Critical Nodes in Interoute Network as Identified by $NVE(v)$.

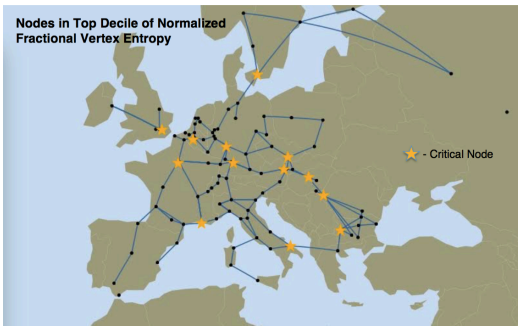


Fig. 10. Critical Nodes in Interoute Network as Identified by $NVE'(v)$.

of over a month of event information from the network, and the subset of those events which were escalated by the customers as incidents. It is evident that events distribute around a peak at 0.175 $NVE'(v)$ whereas, incidents cluster at a peak of 0.95 $NVE'(v)$.

As further justification of the validity of the approach the detailed nature of the data in the Internet Topology Zoo provides the opportunity to see how the local entropy measures are distributed when calculated against real network topologies. In Figures 9 and 10, we highlight against the *Interoute* topology the top 10% of nodes by value of $NVE(v)$ and $NVE'(v)$ respectively. It is striking to note that in both cases these nodes are indeed at critical points in the graph. For example, the nodes with high values occur at points where their removal would cause the creation of a large disconnected component of the graph, and therefore, inflict the highest level of interruption of the operation of the network.

Although the general claims of scale freedom do not fully hold with the real world data we have analyzed in this paper, the motivation to use network structure to identify vulnerable nodes appears promising, and yields two candidates that are locally computable and mirror the behavior of Chromatic and Structural entropy. The justification of studying these values in practical networks has been achieved in theory, and in further work we intend to analyze more real world datasets and extend our entropy measures to include j-spheres for $j > 1$.

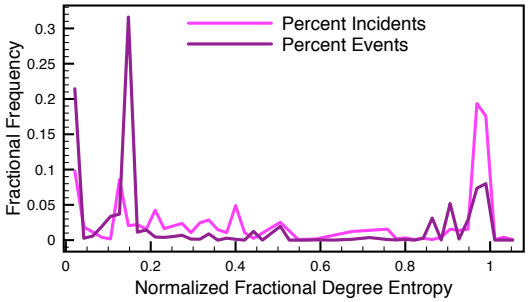


Fig. 11. Distribution of Events and Incidents by $NVE'(v)$ in a real Network of 225,239 Nodes.

REFERENCES

[1] M. L. Steinder and A. S. Sethi, "A survey of fault localization techniques in computer networks," *Science of Computer Programming*, vol. 53, no. 2, pp. 165–194, nov 2004.

[2] M. Miyazawa and K. Nishimura, "Scalable root cause analysis assisted by classified alarm information model based algorithm," in *Proc. of CNSM*, 2011.

[3] C. E. Shannon, "A Mathematical Theory of Communication," *The Bell System Technical Journal*, vol. 27, pp. 379–423, 1948.

[4] G. Simonyi, "Graph entropy: a survey," *Combinatorial Optimization*, vol. 20, pp. 399–441, 1995.

[5] J. Körner, "FredmanKomlós bounds and information theory," pp. 560–570, 1986.

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

[7] M. Faloutsos, P. Faloutsos, and C. Faloutsos, "On Power-Law Relationships of the Internet Topology," *In SIGCOMM*, pp. 251–262, 1999.

[8] L. Li, D. Alderson, W. Willinger, and J. Doyle, "A First-Principles Approach to Understanding the Internet s Router-level Topology," *Acm Sigcomm*, pp. 3–14, 2004.

[9] B. Bollobás and O. Riordan, "Robustness and Vulnerability of Scale-Free Random Graphs," *Internet Mathematics*, vol. 1, no. 1, pp. 1–35, 2004.

[10] B. Bollobás and O. Riordan, "Mathematical results on scale-free random graphs," in *Handbook of Graphs and Networks*. Wiley-VCH, 2006, p. 417.

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

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

[13] W. Willinger, D. Alderson, and J. C. Doyle, "Mathematics and the Internet: A Source of Enormous Confusion and Great Potential," *Notices of the AMS*, vol. 56, no. 5, pp. 586–599, 2009.

[14] 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.

[15] S. Kliger, S. Yemini, and Y. Yemini, "A coding approach to event correlation," ... *Network Management IV*, 1995.

[16] Y. Tang, E. Al-shaer, and K. Joshi, "Reasoning under Uncertainty for Overlay Fault Diagnosis," *IEEE Transactions on Network Service and Management*, vol. 9, no. 1, pp. 34–47, 2012.

[17] C. Borgs, J. Chayes, L. Lovász, V. T. Sós, B. Szegedy, and K. Vesztergombi, "Graph limits and parameter testing," *Proceedings of the thirty-eighth annual ACM symposium on Theory of computing - STOC '06*, p. 261, 2006.

[18] A. Mowshowitz and V. Mitsou, "Entropy, Orbits, and Spectra of Graphs," *Analysis of Complex Networks: From Biology to Linguistics*, pp. 1–22, 2009.

[19] M. Dehmer and A. Mowshowitz, "A history of graph entropy measures," *Information Sciences*, vol. 181, no. 1, pp. 57–78, 2011.

[20] 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.

[21] D. Watts and S. Strogatz, "Collective Dynamics of 'Small-World' Networks," *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.

4.2 Discussion

4.2.1 The General Application of Vertex Entropy as a Root Cause Indicator

As is demonstrated in the following paper presented in Chapter 5, and the discussion that follows in Section 5.2, the vertex entropy measures come close to an effective root cause classifier. However it must be stressed that this is not the intended application of the entropy metrics. The primary motivation for the use of the metrics is to eliminate noise and reduce load on a root cause engine as part of an end to end fault management strategy.

As an example, consider a tree network consisting of a single top of rack switch connecting to N egress switches, each of which has n connections. It is open to debate which of the switches is actually of higher importance to the overall connectivity of the network, and this depends heavily on the overall connectivity. As a matter of fact in this scenario, for $n > N$, inverse degree entropy VE of the top switch is smaller, whereas fractional degree entropy VE' is larger. The normalized variants behave differently as the local topology of each switch is a star network S_n. The modified clustering coefficient of each switch decreases strongly with the degree of the switch node and both normalized variants there increase with increasing degree of the network node. So, in general the switch with the most connections will have a higher entropy metric and be considered more impacting in the topology. Ideally you would want the metric to pass alerts from each of the switches to a system capable of applying downstream logic to such a network, that is the ability to distinguish between alerts on the basis of likely causality due to its position in the networks. Given all of the metrics favor nodes that have a lot of connections and are not locally meshed, it is likely that both of the switches will have a relatively higher value of any of the entropy metrics, compared to, for example, servers connected to the egress switch. In this case, the top of rack switch is likely to be causal if events are present from all of the switches. This further underlines the need for vertex entropy to be used in conjunction with a secondary algorithm to determine root cause.

4.2.2 Post Publication Perspective

This paper was considerably extended in the following publication presented in Chapter 5. A number of the key concepts were treated differently in this publication, due to both the constraints of brevity inherent in a conference proceedings publication, but also due to the evolved understanding of the subject material and the production of many more results from

experiments on the data. in particular, a number of specific statements and claims in the paper I have modified in the subsequent publication. These include:

- **Section III :** In the first paragraph I state that entropy always increases for networks with non-uniform connectivity. In fact, as the analysis shows in the paper, and in particular in the following publication in Chapter 5 this is not strictly true. I prove that the perfect graph maximizes Chromatic Entropy, and minimizes Structural Entropy. The measure of complexity operates in opposite directions for these two entropies and it would have been more strictly correct to have stated so.

- **Section III-A :** The notation in this section has been significantly modified and tightened in the publication presented in Chapter 5. In particular I have adopted the more common notation of k_i for the degree of vertex v_i. In Equation (4) of the paper for example the current notation of $p_i(v_i) = \frac{1}{d_{v_i}}$ is confusing and is written in the following paper as Equation (14) $p_i = \frac{1}{k_i}$.

- **Section III-A :** A number of statements regarding the extremal behavior of the various entropy measures are made without proof or citation. In particular after Equation (7), (11) and (16) claims are made regarding the extremal behavior of the vertex entropy measures. In this paper, due to constraints of brevity, proofs were omitted. In the paper presented in Chapter 5, detailed proofs of these assertions are made, which are consistent with these claims. No available citations were known at the time of publication.

- **Section III-A:** In the discussion after equation (6) I incorrectly write $p_i = C \times \frac{1}{d_{v_i}}$. I should have written $p_i = \frac{1}{C} \times \frac{1}{d_{v_i}}$.

- **Section III:** In Definition 1 of a j-Sphere is inconsistent with how we define the same set in the subsequent paper and how the construct is used to calculate vertex entropies. The inclusion of $j \geq 1$ inside the brackets is not good practice, and the later use of the j-Sphere to capture the central node requires that the distance condition be $d(v_i, v) \leq j$. Definition 1should be more correctly written as it is in Chapter 2, Definition 8.

Chapter 5

Vertex Entropy as a Critical Node Measure in Network Monitoring

5.1 Background to Second Publication

5.1.1 Motivation and Summary of Contribution

The submission to the IEEE journal Transactions on Network and Service Management (TNSM) was motivated as an opportunity to present the full extent of the work on Vertex Entropy that was initially published in the AnNet workshop. Due to the length constraints of the conference format, the original publication did not have the scope to include much of the theoretical work that underpinned the choice of vertex entropy metrics. From an experimental perspective, the results in the first paper are also significantly expanded in this publication and include a methodology to select appropriate values of the entropy metrics to use as a threshold for discarding events.

This submission also benefited from the rejection of a full conference track paper that was submitted to INFOCOM 2017. The reviewers comments are substantially addressed in the manuscript presented to TNSM. After the AnNet publication I was invited to submit an article to the International Journal of Network Management. A manuscript was prepared and submitted, but I subsequently withdrew the paper in favor of this submission.

5.1.2 Theoretical Contribution

A key motivation for the initial research was to identify a viable measure of the contribution of an individual vertex to the entropy of a graph. I significantly built upon the formalism

developed by Dehmer in [23] and [49]. To be able to understand the viability of the proposed forms of entropy it is necessary to establish whether the vertex entropies, when summed across the whole graph, result in an entropy value that is both admissible as an entropy measure and has similar extremal behavior to the more well understood measures.

To be admissible as an entropy metric there are four criteria outlined in [63] and [50], namely *Maximality*, *Additivity*, *Positivity* and *Symmetry*. For each of the proposed metrics a proof is presented of the compliance of the proposed vertex metrics to each of these criteria.

To compare extremal behavior, we first observe that we are only concerned with undirected, connected graphs (see 2), and then establish and prove results demonstrating which networks maximize and minimize the two chosen measures of global graph entropy. The two global measures compared against were structural entropy as defined by Körner in [41], and Chromatic Entropy (this and structural entropy are described in detail in [50]).

The text pays special attention to a small number of prototypal 'special' graphs on n nodes, S_n, K_n, P_n and C_n (see Chapter 2). These special graphs are commonly encountered in communications networks, with K_n representing highly redundantly wired networks and S_n found in aggregation and access networks. The path (P_n) and cycle (C_n) are found in more legacy technologies for local area networks such as token ring and ethernet.

In general Chromatic Entropy is maximized by K_n and minimized by S_n, and structural entropy operates in the opposite way. The paper contains extremal proofs and calculations of extrema for each of these special graphs, both for the proposed vertex entropies and the global entropies, which I believe are novel. These calculations allow identification of which proposed vertex measures behave like structural entropy and which behave like chromatic entropy.

5.1.3 Data and Methods Used

The results presented in this paper were obtained from the same experimental software analysis tools used to generate the results for the first paper. Following comments from the INFOCOM submission however two new vertex entropy measures were added using the clustering coefficient of a node as a direct probability functional following the theoretical formalism of Dehmer.

In addition, statistical tests of correlation were applied to the data using a 2-sample Kolmogorov-Smirnov goodness of fit on the cumulative incident and event distributions [30], to disprove the null hypothesis that the incident and event distributions by vertex entropy are

in fact only different by chance. In each case the null hypothesis could be rejected with a greater than 95% confidence.

The paper also presents analysis using a modified F_β score to identify an ideal value of vertex entropy to use as a threshold above which events are processed. This modified metric balances recall and precision (recall defined as the percentage of incidents at a given value of entropy processed, precision the percentage of incident producing events versus non incident producing events) to bias heavily towards recall. This reflects the fact that the operational usage of vertex entropy is to pre-process events for a root cause detection system in order to reduce event rate load.

5.1.4 Contributions from Co-Authors

As in Chapter 4, the work in this paper is substantially my own, but my co-authors contributed significantly to the development of the ideas, approach and methodology used to analyze the data. In particular the mathematical treatment benefited from a thorough review with my co-authors prior to submission and many hints at data analysis methods (for example the null hypothesis testing) was their suggestion. In addition the manuscript for the paper was prepared and refined with their editorial help.

5.1.5 Related Work

As the paper was an extension to the AnNet workshop paper, the related work is substantially the same. The approach of calculating a graphical entropy as an alert conditioning metric is novel, the closest approach being that of Stearley *et al* described in [66],[67], and [54]

Vertex Entropy as a Critical Node Measure in Network Monitoring

Philip Tee
Moogsoft Inc
1265 Battery Street, San Francisco, CA 94111
phil@moogsoft.com
George Parisis and Ian Wakeman
School of Engineering and Informatics, University of Sussex
Brighton, UK
{g.parisis, i.j.wakeman}@sussex.ac.uk

***Abstract*—Understanding which node failures in a network have more impact is an important problem. Current understanding, motivated by the scale free models of network growth, places emphasis on the degree of the node. This is not a satisfactory measure; the number of connections a node has does not capture how redundantly it is connected into the whole network. Conversely, the structural entropy of a graph captures the resilience of a network well, but is expensive to compute, and, being a global measure, does not attribute any specific value to a given node. This lack of locality prevents the use of global measures as a way of identifying critical nodes. In this paper we introduce local vertex measures of entropy which do not suffer from such drawbacks. In our theoretical analysis we establish the possibility that our local vertex measures approximate global entropy, with the advantage of locality and ease of computation. We establish properties that vertex entropy must have in order to be useful for identifying critical nodes. We have access to a proprietary event, topology and incident dataset from a large commercial network. Using this dataset, we demonstrate a strong correlation between vertex entropy and incident generation over events.**

***Index Terms*—Computer Network Management, Network topology, Network theory, Graph Theory, Entropy.**

I. Introduction and Related Work

Network fault management is principally concerned with the analysis of notifications or events (log messages, SNMP traps etc.) from network devices, with the goal of identifying failures in critical nodes before service is impacted. Events often occur at a very high rate, ranging from 10^2 to 10^6 events per second (eps). In most cases they do not directly indicate a problem. To illustrate, at a typical large enterprise network[1] the event rate is 135 million events a day, whereas there are just a few hundred 'actionable incidents'. The reason for this disparity between the volume of events, and the number of incidents is the over-instrumentation of monitored systems, and the tendency to collect every event for post incident analysis, in case a cause is missed. It is important to state that this heavy event load can render current algorithms used to surface important events unusable, and in many cases operational networks rely upon users reporting failures.

For the purposes of this work we define an event and an incident as follows:

- *Event*: An event is typically a single log message or notification from an underlying monitoring system. We require that it has a timestamp, topology node identifier and description. It is not necessarily a notification of a fault condition, but fault conditions will generally send out at least one event.
- *Incident*: An incident is a support ticket raised as a result of receiving an event, and each incident references a topology node from which the event was received. Although not all incidents are indicative of a significant impact, they are an indication that the node has a fault condition that requires investigation. Typically an incident ticket is raised manually by a support person, or automatically from a monitoring system. Incidents can reference one or more events.

In this paper, we base our analysis on a very large real world network delivering global internet services. More specifically, we have access to the following data[2]:

- **Topology**. The topology is a combination of automatically discovered and manually created datasets. It is normally an example of a Multiplex network, as described in [1]. The analysis presented in this paper is for a network of $225,239$ nodes.
- **Events**. Gathered from the same network is a collection of network events that were monitored over a period of several weeks. For the topology above we analyze $96,325,275$ events.
- **Incidents**. For the same period in which the events were collected, this resulted in $37,099$ such incidents being raised, which in turn refer back to the source event.

Identifying which events are the cause of actual outages is called Root Cause Analysis (RCA) [2]. Many algorithms are used to perform RCA (for a detailed example see [3]), but *scalability limitations* make applying these algorithms to the full event stream impractical. In many cases the maximum event throughput of such algorithms is of the order of 10^2 to

[1]This work is motivated by the experience gained deploying network management software at large commercial scale.

[2]The source of the data is currently confidential, but we are working towards permission to release this dataset with appropriate anonymisation.

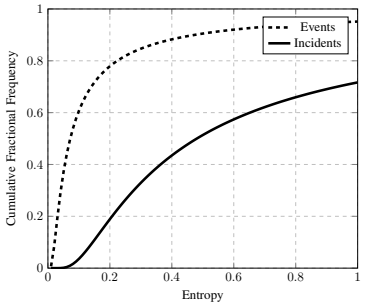


Fig. 1: Ideal Thresholding Cumulative Distribution of Incidents and Events

10^3 events per second (eps). In the example described above the average event rate is 1562 eps, but, from our experience, this can peak to 20,000 eps. To perform RCA across all events, the flow of events has to be significantly reduced, by many orders of magnitude (for example see [4]). Even commonly known techniques, such as compressing repeat events with de-duplication techniques (as described in [5]), which can result in a reduction of events by a factor of 10-100 are not sufficient when event throughputs are often measured in terms of billions of events per day. Failure to control this event rate overload is a principal cause of service outages going undetected by monitoring tools.

The most common approach to reducing the event rate is the simple act of removing uninteresting events with a manual filter or exclusion list, a process known as 'blacklisting'. Blacklisting , which originated in the security event monitoring discipline [6], is extremely time consuming and error prone. At industrial scale, blacklisting can require thousands of rules; in a fast changing network, such an approach is not practical. It is also extremely easy to accidentally blacklist a critical node and miss an event which leads to a service impacting outage. A method to automatically eliminate uninteresting events would yield significant savings, and is the central goal of our research. In particular, we seek a method which can take the topology of a network and automatically discard uninteresting events. The central difference in such an approach from blacklisting is that due to the efficient computability of the metrics discussed in this paper the method can be used even with a dynamic network topology. Blacklisting, by design, is static and requires human intervention to adapt to network changes. Although it may seem potentially risky to throw away events, admitting the possibility that causal events are discarded along with unimportant noise, the alternative is being unable to monitor *any* events and therefore missing *every* causal event.

A. Characteristics of an Ideal Metric

An effective metric should able to identify which nodes are more likely to produce events that will escalate into incidents. An ideal result, given that in the example above only 0.0003% of events get escalated into alerts, would be a metric that can discard 99.999% of events, whilst retaining the few that become incidents. Practically though, given that the goal is to fix the scalability limitations of RCA, we are seeking a metric that can reduce the load by 90%. Further for this to be a practical approach, calculating the metric must itself not present scalability challenges. An ideal metric must:

1) identify which nodes are most likely to produce an incident.
2) allow the discarding of at least 90% of events by the network topology alone.
3) be easy to calculate (not involve any intrinsically non-scalable computational steps) from the network topology alone.
4) be easy to update when the topology changes, ideally involving only computations for a small number of nodes in the region of the network where changes occurred.
5) Assuming a uniform probability of a node emitting an event[3], the metric must clearly segregate a small subset of critical nodes.

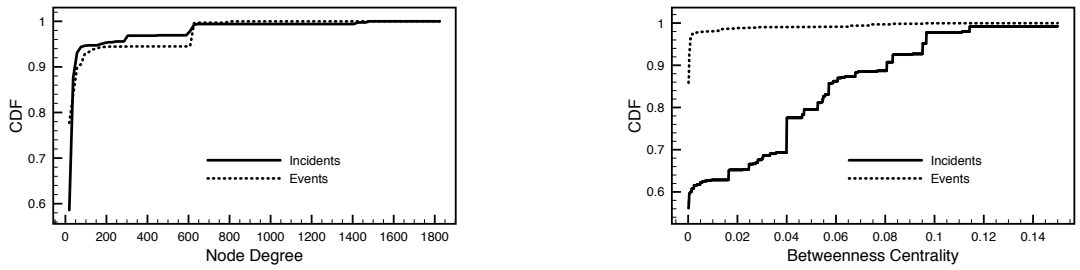
Ultimately the measure of RCA is its ability to capture all root causes and not mis-identify any false positives. This is best described in the language of machine learning using precision and recall. In particular the F_1 score (see [7] for a good description), is a popular measure of the effectiveness of a categorization algorithm such as RCA. Any method which discards root causes (false negatives) along with uninteresting events (true negatives) (or conversely any method that flags root causes (true positives) along with uninteresting events (false positives)) will affect the F_1 score of the overall system. The F_1 metric is most usually defined as the harmonic mean of precision and recall, which we define in Equation (1). In our context $precision$ is measured as the fraction of incidents in the events remaining after discarding all events and incidents that occur below a given value of our metric. Similarly, $recall$ is the fraction of incidents remaining after this discard over all recorded incidents. The value of β in this equation, when set to 1, recovers the standard F_1 measure. In essence when precision and recall are balanced, F_1 is maximized. For our purposes we set a value of β higher to bias the importance of recall over precision in monitoring applications.

$$F_\beta = (1+\beta^2) \times \frac{precision \times recall}{\beta^2 \times precision + recall} \qquad (1)$$

In Figure 1 we illustrate an idealized cumulative distribution of events versus incidents for an ideal metric. This distribution would be achieved if incidents were more likely to occur on nodes with high values of the metric, versus events, according to a distribution around a distinct mean value. This type of skew of incidents towards a higher metric value would allow us to discard events below a given threshold that would remove proportionately far more events than incidents.

A starting place to identify a workable metric is the work of Barabàsi and Albert [8] on network resilience, which was based upon data described by Faloutsos et al [9] and Li et al [10]. Analysis of this data was used by Barabási and Albert to assert that communications networks have a power law

[3]Experience from commercial deployments points to this assumption being reasonable.



(a) (b)

Fig. 2: Cumulative Distribution of Incidents and Events by Node Degree (a) and Betweenness Centrality (b)

node degree sequence, possessing the *Scale-Free* property, whereby, node degree distributions obey the inverse power distribution law. This was further used to justify the claim that communications networks, like the Internet, are both robust to random attack and vulnerable to targeted attack (the central arguments are outlined in [11], [12], [13], and again in [8]). In essence, when removing nodes from a graph randomly, the collapse in connectivity, as measured by the reduction in the size of the giant component, is gradual. However, if nodes are removed by choosing those with highest degree, this reduction is much more rapid (typically removing less than 10% of the nodes will reduce the size of the giant component by more than 90% [8]). It is therefore natural to postulate that node degree could be a metric that satisfies our criteria.

In Figure 2a we present the cumulative distribution of events and incidents by node degree. When we inspect the distribution in Figure 2a the lack of distinction between the cumulative event and incident distribution makes it clear that this does not conform to the idealized distribution in Figure 1. This distinction, which is apparent in the idealized distribution described above, means, at a fixed value of the metric, a far larger proportion of events initiate from nodes of values below this point than incidents. The absence of this preferential tendency for high degree nodes to produce incidents, means degree is a poor metric to achieve a suitable cutoff that would preferentially discard events over incidents. Although degree is extremely easy to calculate (3^{rd} criterion), it fails the first and most important criterion, as it does not provide any useful way of identifying nodes more likely to produce an incident. This lack of correlation is most likely due to high degree nodes being redundantly connected into the network and they may also not impact network function when they fail.

Beyond degree measures there are many other proposed metrics that measure node importance, often centering around centrality measures such as betweenness and eigenvalue centrality ([14], [15]). In Figure 2b we plot the cumulative distribution of events and incidents by betweenness centrality, for our sample data. Betweenness centrality measures the number of shortest paths between any two points in the network that pass by a given node as a fraction of all shortest paths. High values of centrality indicate a node that is critical to the connectivity of the network. It is clear that the effectiveness of this metric is far higher than degree, which is unsurprising as centrality quantifies the importance of a node in terms of connectivity between all points in the graph. Unfortunately the calculation of betweenness centrality scales badly. In the best case for betweenness centrality the fastest known algorithm developed by Brandes ([16], still scales as $O(|V| \times |E|)$, which in the case of our proprietary data is practically unfeasible to compute. To illustrate the problem, on our sample data this calculating the centrality for every node in our proprietary data set required 41 days on server grade hardware. This compares with the entropy metrics described in Section IV, which in identical conditions, require around 1.5 hours to compute every metric for every node sequentially. As our metrics only depend upon local properties of a node and could be calculated locally without a whole graph computation. In practice this means that for a given node, our most efficient metrics VE and VE', compute in less than a second, opening up the possibility that they can be maintained automatically in even the most dynamic environments.

The focus of our research has been with graph entropy, building on the entropy metric presented by Tee et al in [17]. Entropy has been studied in other contexts for anomaly detection (recently [18], and [19] applied the approach to traffic anomaly detection), but graph entropy has received little attention in the context of fault management. As a measure of graph structure it has serious computational drawbacks as its calculation is well known to be *NP-Hard* ([20]), which may account for this. However if these could be overcome with a node level, vertex, approximation, it would be ideal. Using such a node level measure of graph entropy, the proposed technique would be automatically driven from a graph representation of the topology of the monitored network, and importantly could be quickly computed from available inventory databases. Ideally, such a metric would conform to the cumulative distribution illustrated in Figure 1. In this way, at the expense of missing a small number of incidents, the volume of events that need processing can be significantly reduced. As all incidents have an associated event, it is not expected that the distribution would allow perfect recall of incidents as you discard events, but any actual distribution approximating this would be useful in establishing an entropy threshold to allow the discarding of events from nodes less likely to produce an incident. We will seek to demonstrate that our proposed vertex entropy metrics approximate this distribution. A central objective of our research has been to identify easily computable metrics that measure the contribution of an individual node to the entropy of the whole graph. Additionally, for these metrics to be valid entropy

measures, we need to establish their extremal behavior satisfies the criteria of *maximality*, and demonstrate that they satisfy the other essential entropic properties of *additivity*, *symmetry* and *positivity* [21], [22]. Ideally the extremal values of our local variants would coincide with the global entropy measures and provide confidence that these metrics measure the complexity, and therefore, resilience of the networks they represent.

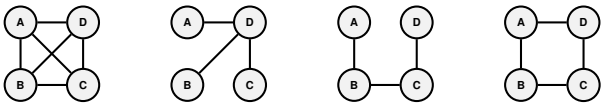
B. Overview

In this paper we describe both the theoretical approach for choosing a valid vertex entropy measure, and also analyze the results when this is applied to our "ground truth" data. Our core motivation is to identify an approximate way of measuring the contribution an individual node makes to the whole graph's entropy, and use that as our metric to eliminate noisy events. However, traditional definitions of graph entropy have insurmountable computational difficulties when applied to networks at scale. The starting point for our investigation is to establish whether there exists node or vertex level measures that when summed across the whole graph behave like the traditional measures. Establishing the existence of such a vertex level metric necessitates an exploration of the characteristics of global entropy measures on simple connected graphs. In section II we present an overview of graph entropy, introducing both *Chromatic* and *Structural Entropies*. Structural Graph Entropy quantifies the degree of connectivity resilience of a graph to edge removal, with low values of structural entropy corresponding to a fully connected or perfect graph, and high values a non-uniform graph with low resilience to edge removal. Chromatic Graph Entropy operates in the reverse sense, with uniform graphs having high chromatic entropy.

A valid entropy measure must satisfy the criteria of *maximality*, *additivity*, *symmetry* and *positivity*. Although *additivity*, *symmetry* and *positivity* are satisfied trivially by the definitions of global entropy, *maximality* is investigated in detail in Section III. We only concern ourselves with simple connected graphs and we prove that for an arbitrary sized graph, the *Star Graph* (S_n) and the *Complete Graph* (K_n) are extremal for both Structural and Chromatic Entropies. Ideally these properties should be shared by our vertex level metrics when summed across the whole graph.

A framework for the construction of node level entropies has been extensively explored in the work of Dehmer *et al*, and summarized in [23]. In section IV we build upon this framework to introduce our proposed forms of local vertex entropy, and investigate their extremal behavior. An important result of our paper is that the vertex entropies we propose have strong analogous behavior to the global variants, when summed across the whole graph, and satisfy *maximality*, *additivity*, *symmetry* and *positivity*. We further demonstrate that our metrics share similar extremal behavior to both global variants.

In section V we evaluate the proposed measures over a large enterprise network. The principal result of our paper is that the vertex entropy measures provide *a computable and effective way to identify important nodes that are more likely to produce incidents*. This is established by identifying that the distribution of event and incident frequency by vertex entropy strongly favors incident production at high values of the metric, verified by analysis of the data using a 2 sample Kolmogorov-Smirnov null hypothesis test to identify whether the distribution of incidents and events by our metrics are trivially correlated. In all cases we can dismiss the null hypothesis and conclude that the metrics produce independent distributions of the events versus the incidents. In addition we calculate for the data a version of the F_1 score, adjusted to account for the preponderance of raw events. Again all proposed metrics demonstrate acceptable improvement in the pre-conditioning of the event data. It is certainly not the case that all incidents occur above a fixed threshold, but at the cost of missing 20% of the incidents, 60-70% of the events can be safely ignored. We conclude our paper in Section VI, with an outlook regarding further research directions.



(a) K_4 (b) S_4 (c) P_4 (d) C_4

Fig. 3: Special Graphs on Four Nodes

II. Theoretical background

Historically, entropy was defined in Graph Theory as a measure of the complexity and non-uniformity of the global structure of a graph. Its use as an analytical tool in network science has been most studied in the dynamical evolution of network growth (see for example [24] and [25]). As a metric it captures many important characteristics, which are of direct interest in a number of applied fields, including the analysis of failure modes of communication networks (see [3]). In particular, networks with non-uniform connectivity will have high values of entropy. Unfortunately the most well understood measures of entropy involve calculations that have impractical computational complexity, as a graph scales in size (see [20] for a good explanation of this point). Further, any change to either the edges or vertices of a graph requires recomputing entropy across the whole graph. It is also extremely difficult to compute the contribution of each individual node to the graph entropy. The variants of Graph Entropy that we explore in this paper are, Körner or Structural Entropy and Chromatic Entropy. Structural entropy measures the mutual information of the stable sets of vertices defined on a graph, a string proxy for the complexity of the graph. Chromatic entropy is defined using the size of subsets of non adjacent vertices, or colorings, of a graph.

In our treatment we confine ourselves to simple, undirected graphs that are connected, and make reference to a number of special graphs, which we define as follows:

- **The Complete Graph** (K_n): This graph is formed from a set of n vertices, maximally connected.
- **The Star Graph on** n **Vertices** (S_n): This graph has one vertex v which is connected to all other vertices, with no other edges in the graph.
- **The Path on** n **Vertices** (P_n): This graph is a simple chain of n vertices, connected by a single edge with no

loops. The path has a single start node v_1 and end node v_n.

- **The Cycle on n Vertices (C_n)**: This graph is a special case of P_n such that $v_1 = v_n$; each node has degree 2.

In Figure 3 we present simple examples of these special graphs with $n = 4$.

Any valid entropy measure must satisfy a number of criteria, (for detailed descriptions see [21], [22]) to be admissible as a well behaved entropy metric. We define these properties on the entropy H of two graphs $F(V, E)$ and $G(V, E)$ as follows:

Definition 1. *For all graphs $F(V, E)$ and $G(V, E')$, sharing the same vertex set V a valid entropy $H(G)$ must satisfy:*

1) ***Additivity**: $H(F \cup G) \leq H(F) + H(G)$*
2) ***Symmetry**: $H(F \cup G) = H(G \cup F)$*
3) ***Positivity**: $\forall G, H(G) \geq 0$*
4) ***Maximality**: For a given collection of vertices V there is an edge set E such that the entropy $H(G)$ of a graph $G(V, E)$ is maximized*

As we explore our candidate entropy measures we will seek to prove that they satisfy these criteria where proofs do not exist in the standard literature.

III. Extremal Behavior of Global Graph Measures

A. Chromatic Entropy

A proper coloring of a graph is the division of the set of vertices V into a collection of subsets such that no member of any subset is adjacent to another member of the same subset.

For a given graph G there maybe multiple colorings, which amount to a collection, or set, of subsets of V. Each of these subsets we call a *Chromatic Class* C_i, with the constraint that $\bigcup_i C_i = V$. The Chromatic Number of a graph, $\chi(G)$, is the smallest number of such subsets that satisfy this constraint. The chromatic number of an graph is bounded by the maximum vertex degree k_{max} [26], [27]:

$$1 \leq \chi(G) \leq (1 + k_{max}) \tag{2}$$

Definition 2. *Chromatic Entropy*

$$I_c(G) = \min \left[-\sum_{C_i} \frac{|C_i|}{n} \log_2 \left(\frac{|C_i|}{n} \right) \right], \forall C_i. \tag{3}$$

where the minimization is over all possible colorings of the graph, and the summation is over all chromatic classes C_i, for a given coloring.

It is possible to establish the following limit on the value of $I_C(G)$:

Theorem 1. *For all graphs G, the Chromatic Entropy $I_c(G)$ is bounded by:*

$$0 \leq I_c(G) \leq \log_2(n) \tag{4}$$

Proof. We note that the lower bound is trivial, and consider the upper bound. We need only to maximize the function $f(p_i) = -\sum_i p_i \log_2(p_i)$ (in our case $p_i = \frac{|C|}{n}$), subject to the constraint $\sum_i p_i = 1$ and $p_i \leq 1, \forall i$, with equality only in the case of a trivial graph of one vertex. Given the definition of $I_C(G)$ as the minimum of equation (3) over all possible colorings, our maximum value will always be an upper bound of I_C. To maximize, we use the method of Lagrange multipliers, considering the following construct, subject to the unity sum constraint $\sum_i p_i = 1$ where $p_i = \frac{|C_i|}{n}$:

$$\mathcal{L} = \max_{p_i} \left[-\sum_i p_i \log_2 p_i - (\lambda - 1) \left(\sum_i p_i - 1 \right) \right] \tag{5}$$

Differentiating by p_i and setting to zero we obtain:

$$\frac{\partial \mathcal{L}}{\partial p_i} = 0; \implies \left(p_i = 2^{(1 - \lambda - \frac{1}{ln(2)})} \right) \forall i \tag{6}$$

From equation (6) our maximum is achieved when all values of p_i are identical and constant. In this case each chromatic class C_i is of identical size $|C_i| = \frac{n}{\chi(G)}$. Feeding this back into equation (3), and substituting for the bounds on $\chi(G)$ from (2) we obtain the desired result.

$$0 \leq I_c(G) \leq \log_2(n)$$

□

In practice these extremal values for $I_c(G)$ are achieved by the perfect graph on n vertices K_n for the maximum, which has a Chromatic Entropy of $\log_2(n)$, and its complement $\overline{K}_n$, where the set of edges is empty, has the minimum value of zero. However $\overline{K}_n$ is not a connected graph; for connected graphs we make the following proposition.

Proposition 1. *For all connected, simple graphs $G(V, E)$ of order $n > 3$ it holds that S_n minimizes $I_c(G)$*

Proof. For $n > 3$ any graph G of n vertices, can be created by progressively adding edges to either S_n or P_n, and by inspection of Table I, S_n has lower entropy than P_n. We will prove our proposition if we can demonstrate that the addition of an edge to any connected graph increases its chromatic entropy, as all graphs obtainable from S_n would have higher chromatic entropy than S_n. Consider any star graph S_n for $n > 3$. If any edge is removed, S_n will cease to be connected, and so by definition is not under consideration of the proposition. As we add edges to the graph S_n the change in chromatic number $\delta(\chi(G))$, can only ever be ≥ 1, or 0. So to complete the proof we consider both cases upon addition of an edge:

Case 1, $\delta(\chi(G)) \geq 1$ ***:*** The addition of a single edge creates an adjacency between two nodes, which must previously have been in the same chromatic class as $\delta(\chi(G)) \geq 1$. If the vertices where not in the same class we cover this in ***Case 2***. The recoloring of the graph will take one or both of the vertices connected by the new edge and add to, or create, a chromatic class of size x. This will reduce the size of a prior chromatic class of size y by x. Edge addition operations that increase chromatic number will always produce classes of increasingly uniform size as we approach a perfect graph K_n. Without loss of generality we will assume that $y > x$, as classes that grow to uniformity will *by necessity* borrow from larger classes as the size of all classes tend to unity. The change in chromatic information due to this re-assignment is:

$$\delta I_c(G) = \frac{x}{n} \log_2 \left(\frac{n}{x} \right) - \left(\frac{y}{n} \log_2 \left(\frac{n}{y} \right) - \frac{y - x}{n} \log_2 \left(\frac{n}{y - x} \right) \right)$$

We seek to prove that $\delta I_c(G) \geq 0$ for all x, y where $y > x$. Elementary manipulation yields the following inequality:

$$\delta I_c(G) \geq 0 \rightarrow x \log_2 \left(\frac{y}{x} - 1 \right) \geq y \log_2 \left(1 - \frac{x}{y} \right)$$

As $\frac{y}{x} - 1 > 1 - \frac{x}{y}$ when $y > x$, we conclude that the inequality holds and $\delta(I_c(G)) \geq 0, \forall n > 3$ under the operation of edge addition when $\delta(\chi(G)) \geq 1$.

Case 2, $\delta(\chi(G)) = 0$ ***:*** **In this instance the addition of an edge does not increase the chromatic number. and as no chromatic classes need to change size,** $\delta I_c(G) = 0$**.**

Eventually additional edges increase the number of adjacencies and consequentially the chromatic number of the graph to its maximum of n, until we arrive at the complete graph K_n, which maximizes $I_c(G)$.In all cases we have seen that adding edges creates a $\delta I_c(G) \geq 0$, and as the first additional edge must belong to ***Case 1***, the proposition is proved.

□

B. Structural Entropy

The original paper of Körner [28], [21] introduced the entropy of graphs by extending traditional Shannon informational entropy. Körner's analysis considered an alphabet of signals, emitted according to a probability distribution, with not all of the alphabet being distinguishable. A graph is constructed such that each member of the alphabet is considered a vertex, with two vertices being connected by an edge if they are distinguishable, and a probability of emission, $P(V)$, being associated with each vertex. To develop the mathematical formulation of structural entropy, Körner introduces a probability distribution $P(V)$, to the normal construct of a graph $G(V, E)$, and defines S to be the maximal set of stable sets of $G(V, E)$. A stable set is a subset of the vertices which are not adjacent to any other member of the stable set, the maximal set being the collection of largest stable sets.. A number of equivalent definitions of structural entropy, $H(G, P)$ are possible, of which the simplest is in terms of the mutual information between $P(V)$ and $G(V, E)$ as follows [21]

Definition 3. *Körner or Structural Graph Entropy*

$$H(G, P) = H(P) - H(P|S) \tag{7}$$

This measure, which we call structural entropy, is related closely to the Chromatic Entropy. In our treatment we identify $P(V)$ with the probability of the emission of an event, which we further assume to be uniform. With that simplification the two quantities are related as follows (for an in depth treatment see [22]):

$$H(G, P) = \log_2(n) - I_c(G) \tag{8}$$

Structural entropy can most easily be interpreted as quantifying the extent to which the local neighborhood of a node is unique. In other words the value of $H(G, P)$ is minimized when all vertices are equivalently connected, and maximized when each node is distinguishable by its local topology. Given equation (8) we can state the following lemma on the bounds for $H(G, P)$.

Lemma 1. *For any graph $G(V, E)$ on n nodes, assuming that P is uniform, the structural graph entropy is bounded as follows:*

$$0 \leq H(G, P) \leq \log_2(n) \tag{9}$$

Proof. This bounding of $H(G, P)$ is easily verified by direct substitution of (4) into (8), and as such the proof is trivial. □

We summarize the extremal behavior of our global graph measures in table II.

TABLE I: Values of Global Entropies for Special Graphs

	$I_c(G)$	$H(G, P)$ - P Uniform
S_n	$\frac{n-1}{n} \log_2 \left(\frac{n}{n-1}\right)$ $-\frac{1}{n} \log_2(n)$	$\frac{n-1}{n} \log_2(n-1))$
K_n	$\log_2(n)$	0
P_n, n even	1	$\log_2(n) - 1$
P_n, n odd	$1 + \log_2(n) - (n+1)\log_2(n+1)$ $-(n-1)\log_2(n-1)$	$(n+1)\log_2(n+1)$ $+(n-1)\log_2(n-1)$ -1
C_n, n even	1	$\log_2(n) - 1$
C_n, n odd	$\log_2(n) - \frac{n-1}{n}\Big(1 - \log_2(n-1)\Big)$	$\frac{1-n}{n}\Big(1 - \log_2(n-1)\Big)$

TABLE II: Graph Types that Maximise and Minimize Entropy

	Chromatic	Structural
Maximum	K_n	S_n
Minimum	S_n	K_n

IV. Local Vertex Entropy Measures

Recent work on Graph Entropy by Dehmer [23], [29] provides a framework that unifies the global invariants discussed, and provides a pathway to extend these measures in a more computable direction. Both Structural and Chromatic entropy rely upon partitions of the vertex set of the graph, which are known *NP-Hard* problems.

Dehmer defines the concept of a *local functional* for a vertex, which can be scoped to calculate values for every vertex based upon the local topology of the graph. The degree of locality in the treatment is controlled by using the concept of *j-spheres*, S^j in the graph, centered at a given vertex. For clarity, in the definition that follows a superscript indicates the order of the j-sphere, whereas subscripts run over the members of the vertex set of the graph. Dehmer's original definition relied upon subsets of vertices of a fixed distance from a given vertex v_i. where distance $d(v_i, v_j)$ is the shortest distance between distinct vertices v_i and v_j (i.e. $i \neq j$). The distance is measured in the number of edges traversed in a walk from v_i to v_j, and in communications networks is commonly referred to as the 'hop' count. This definition excluded the vertex v_i, and other interior nodes for $j \geq 1$, but in our later treatment this introduces problematic zeroes when we define the clustering coefficient. We extend the definition of a j-sphere to include the node v_i as part of the set. This avoids certain special graphs such as S_n having zero clustering coefficients that would

introduce infinities into our later definitions of normalized entropies. This is different to the definition given by Dehmer, in that we include all interior nodes to a given j-Sphere. The definition so modified is as follows:

Definition 4. *For a graph $G(V, E)$, we define for a node $v_i \in V$, the 'j-sphere' centered on v_i as:*

$$S_i^j = \{v_k \in V | d(v_i, v_k) \leq j, j \geq 1\} \cup \{v_i\} \tag{10}$$

and for convenience when we define the clustering coefficient in equation (21), the related 'j-edges' E_i^j as

$$E_i^j = \{e_{kl} \in E | v_k, v_l \in S_i^j\} \tag{11}$$

In essence the sets S_i^j and E_i^j are the local j-hop neighborhood of the node v_i, with S_i^j being the collection of all nodes j hops away from v_i, and E_i^j being the set of edges between them.

The concept of j-spheres is a very convenient formalism to capture *locality* in the graph. Essentially j can range from 1 to the diameter, $D(G)$, of the graph (as defined as the maximum length shortest path between two nodes). By breaking a large graph into j-spheres, we can progressively examine complex combinatorial quantities such as graph entropy on increasingly larger subsets of the graph until at $j = D(G)$ the global value is being effectively computed. Using our extended definition, we proceed by equipping each S_i^j with a positive real-valued function $f_i : v_i \in S_i^j \to \mathbb{R}^+$. This function is proposed to be dependent upon properties of the nodes that are members of the j-sphere, such as their degree, number of cycles and so on, which capture the local structural properties of the graph. From this, we can construct a probability function for each vertex as

$$p_i = \frac{f_i}{\sum_{v_j \in V} f_j} \tag{12}$$

which trivially satisfies $\sum_i p_i = 1$.

Essentially these functions are used to construct entropy measures in direct analogy to Shannon entropy as follows:

$$H(v_i) = -p_i \log_2 p_i \tag{13}$$

The principal direction of Dehmer's proposition is that these functions f_i when used to construct entropy, describe the local 'information' that a given vertex carries about the global structure of the graph. However, in the published work [23], [29], these functions are complex expressions, which introduce global invariants of the graph complicating their computation.

We can now apply Dehmer's formalism using the available invariants available in j-spheres for different values of j. For reasons of computational simplicity in this work we restrict ourselves to $j = 1$, which is the immediate local neighborhood of a given node. Although this sacrifices global structure of the graph, we will show that the results are still of operational significance and, because of locality, very efficient to compute. Indeed if $\langle k \rangle$ is the average degree of a node, most of our metrics are computable in $O(|V| \times \langle k \rangle)$, significantly less than, for example, centrality measures. Given the constraint of locality, a number of constructs can be designed that satisfy the probability functional defined in equation (12) up to a normalization constant. In the immediate neighborhood of a vertex the available measures are restricted to the degree of the vertex k_i, and the presence of cycles in the local subgraph. It is important that the measures that are constructed are bounded in an acceptable way, when summed across the whole graph and satisfy the fundamental properties of an entropy measure: *maximality*, *additivity*, *symmetry* and *positivity* [21], [22].

In Table III we summarize the available probability constructs that we will investigate. For j-spheres where $j > 1$ we have not conducted any analysis, and this remains an open question for further research. It should be noted though that as j approaches $D(G)$, the diameter of the network, the probability functionals approach a constant value, which is unlikely to reveal much of the structure of the network.

TABLE III: Local Probability Functional Constructs on a j-sphere

	$j = 1$	$j > 1$	$j = D(G)$
$\frac{1}{k_i}$	$VE(v)$	Unexplored, $\frac{1}{\lvert E_i^j \rvert}$	Constant Value $\frac{1}{\lvert E \rvert}$
$\frac{k_i}{\lvert E \rvert}$	$VE'(v))$	Unexplored, $\frac{\lvert E_i^j \rvert}{\lvert E \rvert}$	Constant Value 1
C_i^j	$NVE(v)$,$NVE'(v)$, $CE(v)$,$CVE'(v)$	Unexplored	Unexplored

A. Inverse Degree Entropy

The first and most basic probability functional, which we can construct on the 1-sphere of a vertex, uses its inverse degree k_i and is defined as follows:

$$p_i = \frac{1}{k_i} \tag{14}$$

and the corresponding entropy of the vertex $VE(v_i)$, and whole graph $H_{InvDegree}$ as

$$VE(v_i) = \frac{1}{k_i} \log_2(k_i), \tag{15}$$

for the whole graph:

$$H_{VE} = \sum_{i=0}^{i<n} \frac{1}{k_i} \log_2(k_i) \tag{16}$$

The first observation is that the sum of inverse degrees does not satisfy the constraint $\sum_i p_i = 1$. However, one can observe that for any given graph G, this probability functional sums to the constant:

$$C = \sum_{i=0}^{i<n} p_i = \frac{\sum_{i=0}^{i<n} \left(\prod_{j \neq i} k_j \right)}{\prod_{i=0}^{i<n} k_i} \tag{17}$$

We note that $p_i = \frac{1}{C} \times \frac{1}{k_i}$, and discard the constant as part of the normalization.

As the expression in equation (15) involves a sum of logarithmic terms (which are all positive), the conditions of *additivity*, *symmetry* and *positivity* are satisfied trivially, in particular for additivity as the combination of two graphs must as a minimum increase the degree of a vertex from each graph, or leave the degrees of the two graphs unchanged, the combined graphs entropy will be greater than or equal to the

sum of the two graphs, thereby satisfying the additivity criteria of Definition 1.

Regarding *maximality*, it suffices to establish that equation (16) has a maximum for a fixed set of vertices and edges. This can be done using Lagrange multipliers with the constraint $\sum_i p_i = C$, where C is the constant from equation (17). This yields an expression for the p_i as $p_i = 2^{(C-1-\lambda)}$, where λ is the Lagrange multiplier, confirming that the entropy has a maximal value for a graph whose degrees are equal. Referring to Table IV, we can see this is obtained by the cycle graph on n vertices C_n. Indeed the special graphs are ordered by increasing inverse degree entropies in the sequence $S_n < K_n < P_n < C_n$.

TABLE IV: Values of Vertex Entropy for Special Graphs

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

B. Fractional Degree Entropy

Inverse degree is unsatisfactory. Firstly the probability functional is not naturally defined to satisfy the unity sum constraint. Secondly, and more importantly, the degree of a vertex does not capture how '*hub-like*' the node is relative to others. To capture this, we can define an alternative functional, which is based upon the ratio of the vertex degree to the total number of edges in the graph, as follows:

$$p_i = \frac{k_i}{2|E|} \tag{18}$$

Given that $\sum_{v_i \in V} k_i = 2|E|$ this functional directly satisfies the unity sum constraint. In a parallel way to equation (15), we define the fractional degree entropy as:

$$VE'(v_i) = \frac{k_i}{2|E|}\log_2\Big(\frac{2|E|}{k_i}\Big), \tag{19}$$

for the whole graph:

$$H_{VE'} = \sum_{i=0}^{i<n} \frac{k_i}{2|E|}\log_2\Big(\frac{2|E|}{k_i}\Big) \tag{20}$$

Following the treatment of Inverse Degree Entropy, we note that the expression in equation (20) again involves a sum of logarithmic terms (which are all positive), so the conditions of *additivity*, *symmetry* and *positivity* are satisfied. To establish maximality, we can again use the technique of Lagrange multipliers using the constraint $\sum_i p_i = 1$, which yields a similar result to inverse degree entropy that the maximal value is obtained for a graph with equal vertex degrees satisfying $p_i = 2^{1-\lambda}$. In Table IV this is satisfied by K_n and C_n. The special graphs using this measure are ordered in increasing fractional degree entropy as $S_n < P_n < C_n = K_n$. We summarize these results in Table V.

TABLE V: Extremal Graphs for Unnormalized Vertex Entropy

	VE	VE'
Maximum	C_n	$K_n = C_n$
Minimum	S_n	S_n

C. Normalized Degree Entropy

There is a considerable practical difference between a star network topology (S_n) and a fully meshed one (K_n). In the former, the network is vulnerable to the loss of its central high degree vertex; in the latter, the loss of any one vertex can never create isolated vertices. Both prior measures make little distinction between these two topologies for nodes of identical degree, but there are available metrics measurable at one hop distance that capture this concept. Indeed, in the case of fractional degree, there is no way for the degree to capture the intricacies of the local topology of the node. Introduced in [30] and [8] is the concept of the clustering coefficient of a vertex. The traditional definition counts edges between neighbors of a vertex, which yields a zero value for S_n that is problematic in our treatment. We avoid zeros using our extended version of the j-sphere in equation (10). In terms of the degree of vertex i, k_i, the following definition captures how similar the j-sphere surrounding a vertex is to the complete graph K_n and is defined in terms of the 1-sphere edge set E_i^j as:

$$C_i^1 = \frac{2|E_i^j|}{k_i(k_i+1)} \tag{21}$$

In essence the clustering coefficient measures the probability that two randomly chosen nodes in the 1-hop subgraph have an edge between them. In this way the lower the value of the coefficient, the higher the likelihood that the failure of the node at the center of the subgraph will cause two nodes to become disconnected (see for example [31]). This completely captures how well meshed a node is into its local neighborhood, and therefore serves as an ideal candidate for further refining the vertex measures introduced earlier. In particular, we want to highlight vertices whose clustering coefficient is low, that is, their local neighborhood is more similar to S_n locally than K_n. To that end we define the following *Normalized* Vertex Entropies:

Definition 5. *We define for a graph $G(V,E)$ the following* Normalized *Inverse Degree Entropy for both vertex and total graph as follows:*

$$NVE(v_i) = \frac{1}{C_i^1} \times VE(v_i), \tag{22}$$

for the whole graph:

$$H_{NVE} = \sum_{i=0}^{i<n} \frac{(k_i+1)}{2|E_i^1|}\log_2(k_i), \tag{23}$$

and the corresponding definition for fractional vertex entropy is defined similarly:

$$NVE'(v_i) = \frac{1}{C_i^1} \times VE'(v_i), \tag{24}$$

and total entropy:

$$H_{NVE'} = \sum_{i=0}^{i<n} \frac{k_i^2(k_i+1)}{4|E||E^1(v_i)|} \log_2\left(\frac{2|E|}{k_i}\right) \tag{25}$$

Proving compliance with Definition 1 for these normalized values is not as straightforward as the non normalized values. However, as the expression in equation (21) is always positive, the *symmetry* and *positivity* criteria are automatically satisfied. With regard to additivity and criteria 1 of Definition 1, although not a rigorous proof, considering two graphs being minimally joined by a single vertex, the clustering coefficient of that vertex will *decrease* and so the value of NVE or NVE' of the shared vertex will increase, satisfying the inequality.

For maximality, the introduction of the clustering coefficient complicates the use of the Lagrange multiplier method, as p_i and C_i^1 are related quantities. It is the beyond the scope of this work to present a formal proof of *maximality* but we can calculate the values of the normalized entropies for our special graphs and we summarize the results in Tables VI and VII. The special graphs using NVE ordered in increasing entropy are in the sequence S_n, K_n, P_n, C_n and for NVE', K_n, C_n, P_n, S_n. With the assumption that it is possible to maximize these entropies these values are admissible measures of entropy. It is interesting to note that the distinction between star topologies and meshed ones is much less distinct with NVE. Comparing extremal behaviors to our global entropy measures, we identify NVE with Chromatic entropy and NVE' with Structural entropy.

TABLE VI: 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	$\frac{3}{4}n$	$\frac{3}{2}\log_2(n)$

TABLE VII: Maximal and Minimal Total Vertex Entropy Graph Types

	NVE	NVE'
Maximum	C_n	S_n
Minimum	S_n	K_n

D. Alternative Vertex Entropy Constructions

The local clustering coefficient C_i^1 can also be used to construct two alternative probability functionals, which an exhaustive study necessitates. In the first instance, as the clustering coefficient itself is a value strictly in the range $(0, 1]$ it is a valid informational functional in its own right. We can define a clustering coefficient entropy, $CE(v_i)$ by identifying $p_i = C_i^1$, as follows:

Definition 6. *For a graph $G(V, E)$ the clustering coefficient entropy, $CE(v_i)$ of a vertex v_i is defined as*

$$CE(v_i) = C_i^1 \log_2 C_i^1, \tag{26}$$

and for the whole graph:

$$H_{CE} = \sum_{i=0}^{i<n} C_i^1 \log_2 C_i^1 \tag{27}$$

In addition, we can also approach the normalization of the fractional vertex entropy by defining an alternative probability functional using the clustering coefficient as:

$$p_i = \frac{1}{C_i^1} \times \frac{k_i}{|E|} \tag{28}$$

This probability functional is within the range $(0, 1]$ as for a given vertex this simplifies to $p_i = \frac{|E_i^1|}{(k_i+1)|E|}$, which for a connected node is strictly non-zero and $|E_i^1| \leq |E|$. It is not possible to extend the inverse degree functional in a similar way as the equivalent definition $p_I = \frac{1}{k_i C_i^1}$ is not bounded to fall into the range $(0, 1]$. We therefore make the following definition for the Cluster Coefficient Fractional Degree Entropy as follows:

Definition 7. *For a graph $G(V, E)$ the Cluster Coefficient Fractional Degree Entropy $CVE'(v_i)$ of a vertex v_i is defined as:*

$$CVE'(v_i) = \frac{k_i}{C_i^1|E|} \log_2\left(\frac{C_i^1|E|}{k_i}\right), \tag{29}$$

and for the whole graph:

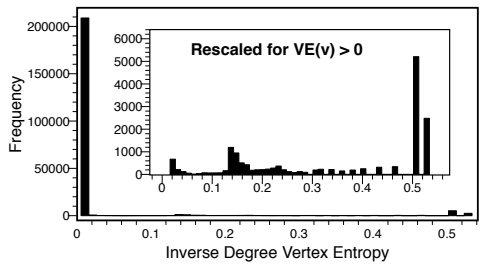
$$H_{CVE'} = \sum_{i=0}^{i<n} \frac{k_i}{C_i^1|E|} \log_2\left(\frac{C_i^1|E|}{k_i}\right), \tag{30}$$

Using similar arguments to the previous entropy types we can establish conformance with *additivity*, *symmetry* and *positivity* of Definition 1 by observing that in equations (27) and (30) are sums of logarithms . The remaining property of *maximality*, in complex to verify due to similar issues to the normalized entropy values NVE and NVE'. It is beyond the scope of this paper to present a rigorous proof of *maximality*, but we can calculate the values for our special graphs, which we summarize in Table VIII. The special graphs using CE ordered by increasing entropy are in the sequence K_n, S_n, P_n, C_n and for CVE', S_n, C_n, P_n, K_n.

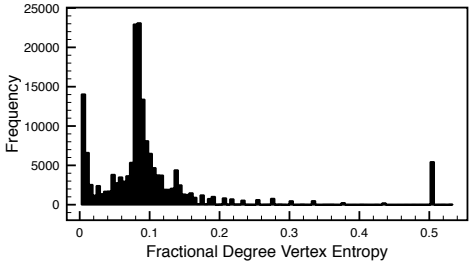
TABLE VIII: Values of Clustering Coefficient Entropies for Special Graphs

	CE	CVE'
S_n	$\frac{2}{n}\log_2(\frac{n}{2})$	$\log_2(n-1) - \frac{n}{2}\log_2(n)$
K_n	0	$2\log_2(\frac{n}{2})$
P_n	$\frac{2(n-2)}{3}\log_2(\frac{3}{2})$	$\frac{1}{n-1}\left[3(n-2)\log_2(\frac{n-1}{3}) - 2\log_2(n-1)\right]$
C_n	$\frac{2n}{3}\log_2(\frac{3}{2})$	$3\log_2(\frac{n}{3})$

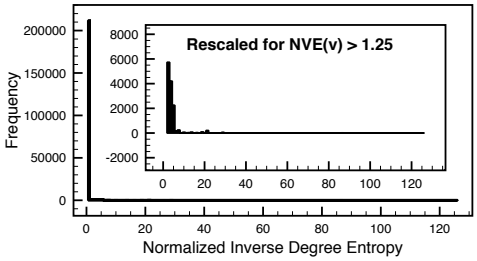
From these calculations we can summarize in Table IX the extremal graphs for these entropies.



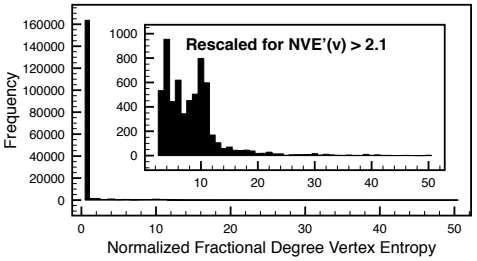
(a) Frequency Distribution of $VE(v)$



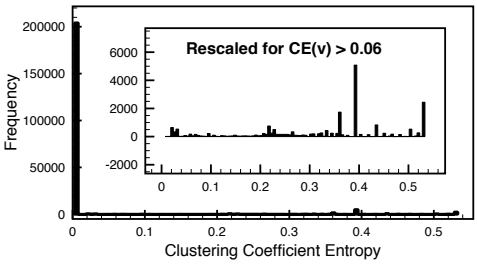
(b) Frequency Distribution of $VE'(v)$



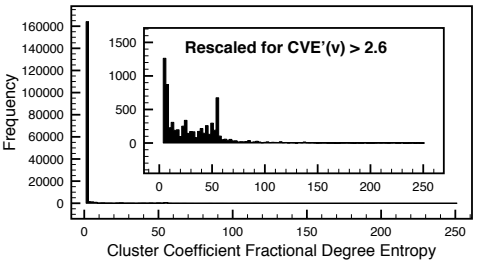
(c) Frequency Distribution of $NVE(v)$



(d) Frequency Distribution of $NVE'(v)$



(e) Frequency Distribution of $CE(v)$



(f) Frequency Distribution of $CVE'(v)$

Fig. 4: Frequency Distributions for a Network of 225,239 Nodes

TABLE IX: Extremal Graphs for Clustering Coefficient Entropy

	CE	CVE'
Maximum	C_n	K_n
Minimum	K_n	S_n

V. Evaluation and Discussion

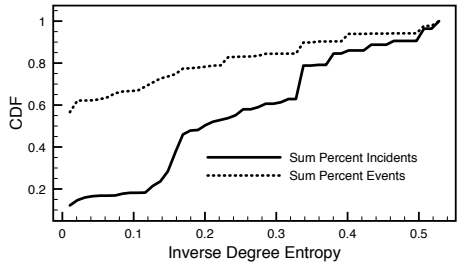
A. Data and Methods

We analyzed data from a large operational dataset obtained from a web portal operator. In previous work [17] we also applied our techniques to the 'Internet Topology Zoo' (ITZ) ([32]), but this critically does not have any event or incident data.

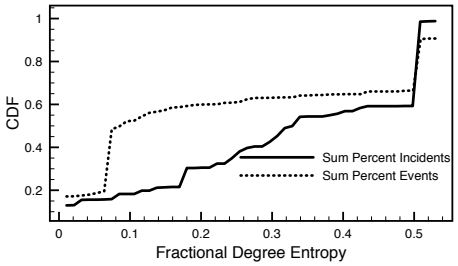
Our commercial data, however, contains a rich source of events and incidents, and in particular allows the analysis of event and incident distribution by originating node. The analysis was performed using a suite of software tools implemented in JAVA, and operated in conjunction with a MySQL database for permanent storage[4]. A brief description is below:

- `graph_analyser`: This executable was built to ingest source topology as a list of edges in a comma separated file format. The program calculates all of the metrics described in Section IV and betweenness centrality and stores the results both in raw and frequency distribution format in the database. The value stored in the database are used by the other analysis programs to produce distributions of events and incidents by node metric.
- `event_analyser`: This executable ingests and parses the full sample of events obtained from the customer. Each event is presented as a string of symbols separated by the '|' character. The format of the events followed a fixed pattern with the syntax: timestamp |datacenter | application | node | description. After each event is parsed the executable populates a distribution of event count by value of the metric for each value of 'node'.
- `incident_analyzer`: This executable operates in an almost identical fashion to the event analyzer, but instead analyzes data that is obtained from a report ran on the customer's incident management system. Each incident represents an event that has been escalated according to their manual triage process and is presented with the following syntax: date | timestamp |datacenter |

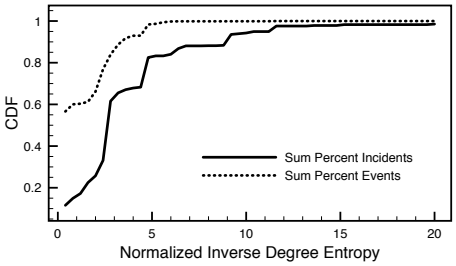
[4]The source code for these analysis tools is available at https://github.com/philtee2001/analyzer.git, and instructions for building are available from phil@moogsoft.com



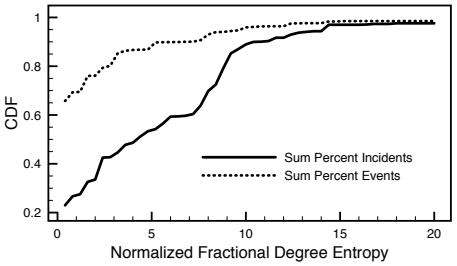
(a) Cumulative Distribution of Events and Incidents by $VE(v)$



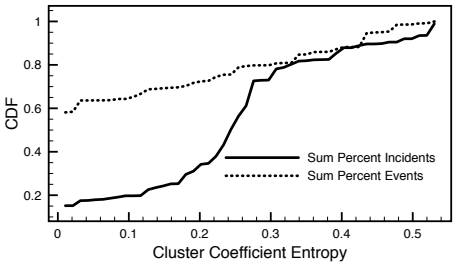
(b) Cumulative Distribution of Events and Incidents by $VE'(v)$



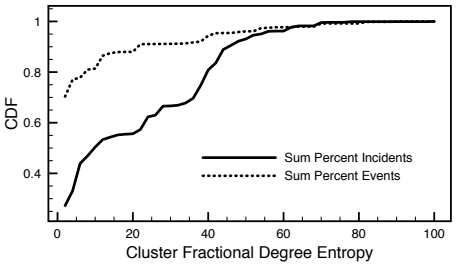
(c) Cumulative Distribution of Events and Incidents by $NVE'(v)$



(d) Cumulative Distribution of Events and Incidents by $NVE'(v)$



(e) Cumulative Distribution of Incidents by $CE(v)$



(f) Cumulative Distribution of Events and Incidents by $CVE'(v)$

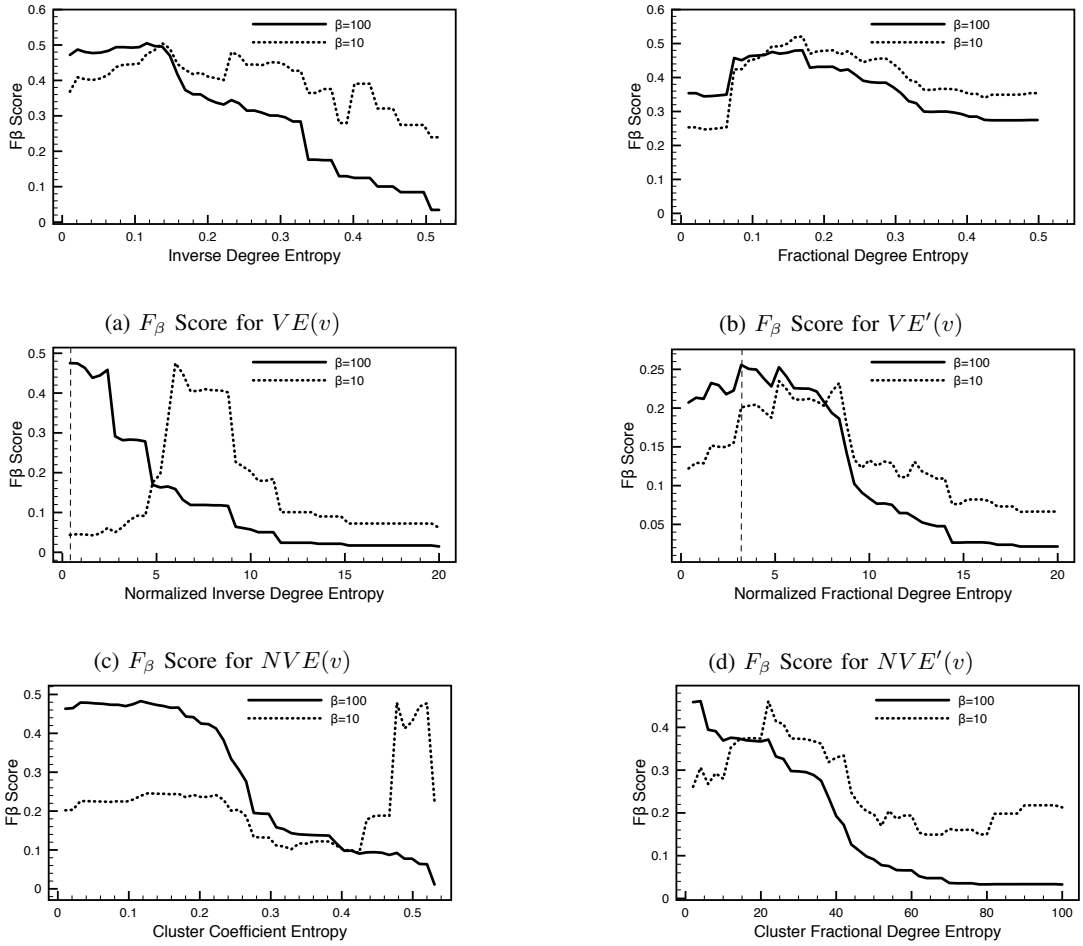
Fig. 5: Cumulative Distributions of Events and Incidents in a Network of 225,239 Nodes

ticket number | node | type | property | agent | description. node field ties directly back to the topology data and is used to populate a distribution of incidents by considered metric. For this data not all values of 'type' are considered, as they indicate whether or not the incident was deemed to be significant. We discard any incidents that were not accepted by the help desk without escalation.

B. Evaluation

Using the dataset described in Section V-A, we begin in Figure 4 by plotting the distribution of nodes by the various entropy measures. For a number of the metrics, the data is heavily skewed by large numbers of the nodes having a zero of low value. In Figures 4a, 4c, 4d, 4e and 4f we plot the distribution excluding these values, rescaled. All of the measures share a common feature in that the vast majority of the nodes posses a heavy skew towards low values of the metrics. This is encouraging, because for an entropy metric to be useful in identifying important nodes a uniform distribution would be unexpected. Except in the case of fractional degree entropy VE' (Figure 4b) the skew is so pronounced that to illustrate the distribution above minimal values of the metric we have embedded a subgraph rescaled to eliminate the dominating cluster of values towards the low values of the metric.

With both inverse degree $VE(v)$ and fractional degree entropy $VE'(v)$ the distribution achieves the first objectives of being non-uniform and separating out a small subset of nodes with high values of the metric. In the earlier discussion in section I, this distribution profile was a necessary condition of the metric having utility when identifying nodes likely to produce incidents. However, these metrics do not distinguish between a high degree node that has many redundant paths into the network and one that does not. In our theoretical analysis in Section III, we identified the need to highlight nodes whose local topology was more similar to S_n than K_n, which the non-normalized metrics do not. The point of our normalized metrics is to capture this aspect of local topology and provide a way of identifying nodes that have high degree but low redundancy. From considerations of network design, these nodes are more likely to produce events that escalate



(a) F_β Score for $VE(v)$

(b) F_β Score for $VE'(v)$

(c) F_β Score for $NVE(v)$

(d) F_β Score for $NVE'(v)$

(e) F_β Score for $CE(v)$

(f) F_β Score for $CVE'(v)$

Fig. 6: F_β Score Plots in a Network of 225,239 Nodes

into incidents when they fail.

To establish whether the data supports this hypothesis, we turn to the distributions of normalized inverse degree $NVE(v)$ and normalized fractional degree entropy $NVE'(v)$ in Figure 4c and Figure 4d. It is interesting to note that both quantities share the same non-uniform distribution as the non-normalized forms, with a much more pronounced separation of the extremal values. This is consistent with our supposition that the normalized metrics exclude a subset of high degree nodes that have multiple paths through the network.

To fully exhaust all potential metrics available on a 1-sphere, we also plot the distributions for clustering coefficient $CE(v)$ and cluster coefficient fractional entropy $CVE'(v)$ in Figures 4e and 4f. Again these distributions are skewed fairly heavily towards low values of the metric, and show interesting, much smaller clusters at higher values of the metrics.

Our central claim is that the local measures of vertex entropy are more effective at identifying nodes that will generate incidents than simply selecting the nodes of highest degree, as suggested by scale free models of dynamic networks. In Figure 2a we presented the distribution of events and incidents by node degree, from which it is clear that there is very little difference in the distribution between events and incidents, and that there are no useful distinctions between the distribution of incidents by degree versus events. Although high degree nodes are more likely to cause impact than low degree nodes when failing, network design usually mitigates failure points by adding in redundant paths through the network to avoid single points of failure. This is further underlined by the cumulative distribution plot in Figure 2a, where it is evident that the distribution of events and incidents is effectively the same. In Table XI we note that the 2-sample Kolmogorov-Smirnov test does not allow us to dismiss the null hypothesis, with a P-Value in excess of the α value, indicating that degree is not a discriminatory factor. A side effect of this analysis is affirmation that one of our key assumptions that events are emitted with uniform probability across all nodes. For these reasons, degree is not a reliable indicator of impact when a node fails.

To contrast this with our entropy based metrics in Figure 5 we plot the cumulative distributions of events and incidents by each of our candidate metrics. In each case there is a heavy skew towards higher values of the metrics for incidents versus events. This is the first indication that the vertex entropy metrics are indeed useful for identifying nodes more likely to produce incidents. As discussed in the introduction we can

make use of the F score methodology to identify how effective our entropy metrics are at optimizing recall and precision. To do so, however, we must build upon the basic measure introduced in [7] to take account of the fact that our metric is being proposed to pre-condition data before a categorization algorithm (RCA) is used to determine whether an event is causal. In general raw events contain many duplicate notifications which can be compressed by the application of a de-duplication (see for example [5]). This can result in the number of events being compressed by a factor of 10-100. In addition, the *cost* of a missed incident is significantly more impactful to a business than the cost of processing an event. As the basic measure of an F score assumes equal weight, we instead adopt a weighting factor β and measure the F_β score for our event and incident distributions. The F_β score is defined in equation (1). Effectively this metric measures the balanced effectiveness of an algorithm at identifying true positives without producing too much noise in the form of false positives. In the context of event management and RCA, this is the ability of an algorithm to capture every incident without surfacing *false* incident notifications. The typical application of the F_β score though weights precision and recall evenly, and given that a missed incident is potentially costly, the β parameter allows us to bias in favor of recall. We choose a heavy bias of $\beta = 100$.

In Figure 6 we plot for each of our metrics the F_β scores as a function of the metric for a $\beta = 10$ and $\beta = 100$. Plotting the F_β score identifies a value of the entropy metric that maximizes the F_β score. This maximum corresponds to the best threshold to use to discard events from nodes with a value of entropy that is below it, allowing you to reduce event load whilst preserving events that are likely to escalate into incidents. These plots illustrate the importance of the weighting factor in the F_β score for identifying the correct choice of entropy to set a discard threshold at. In each case the $\beta = 100$ establishes a lower discard threshold as you would expect, given that we are treating recall as more important than precision, as the maxima of the F_β score occurs at a lower value of the entropy measure. In Table X we collect the discard rates at the maximum of the F_β score for $\beta = 100$. In each case it is evident that it is possible to choose a value of the metric, in this case our choice of vertex entropy, that will selectively discard many more events than incidents, and in fact, by the nature of the scaling of the F_β score, at a value of entropy that would discard 20% of the incidents, some 65% or $15,000,000$ events can be safely discarded. For the data we analyzed , this amounts to discarding $62,600,000$ events *before* expensive RCA processing. This amounts to reducing the event rate from approximately 12 per second to 4, which operationally could be very significant. In order to replicate this result using manual blacklisting, this would require the maintenance of a list of nodes that are relatively unimportant. In the case of the network we analyzed, that would amount to some $200,000$ nodes, which are apt to change frequently. As we indicated in the Section I alternative simpler metrics such as node degree are unable to achieve similar effectiveness in identifying important incident producing nodes as our entropy metrics or centrality measures.

TABLE X: Maximal % Discards of Events and Incidents ($\beta = 100$)

Metric	Max Value	% Events	%Incidents
VE	0.116	87%	52%
VE'	0.170	68%	22%
NVE	0.400	57%	12%
NVE'	3.200	85%	45%
CE	0.127	67%	20%
CVE'	4.000	76%	32%

To further test the correlation between our vertex entropies and incident creation, statistical hypothesis testing of the distributions using a 2-sample Kolmogorov-Smirnov goodness of fit between cumulative distributions of events and incidents was undertaken. Using an α of 5%, and assuming the *Null Hypothesis* that both event and incident distributions of all metrics shared the same cumulative distribution, very low P-Values were obtained, indicating that the difference in distributions is highly unlikely to be the effect of randomness. We summarize the findings in Table XI. This result convincingly contradicts the *Null Hypothesis*, and we can safely conclude the difference in the distribution is a result of a strong correlation between high values of both metrics, and a higher likelihood of events escalating into incidents. This result continues to be valid down to values of $\alpha = 1\%$, and is a strong indication that our local metrics are capturing enough of the local topology of the network to be useful as a way of assessing the impact of a nodes failure on the overall connectivity of the network. In essence, impact is a result of the node being part of a large number of shortest paths between any two arbitrary points in the network. Although high degree makes it more likely, the similarity of the local topology of the node to K_n versus S_n mitigates that, and our normalized metrics successfully account for this subtlety. It is interesting to note that the *Null Hypothesis* cannot be dismissed for the degree distributions as the P-Value is higher than $\alpha = 5\%$.

It is interesting to speculate which of the metrics is the most effective metric to use to pre-condition events for RCA. In practice any of the metrics investigated appear to have merit, but it is important to note that the local clustering coefficient of a node can be expensive to compute for highly connected and nodes in a heavily meshed network. For a network that is maximally connected with n nodes, the calculation of the clustering coefficient is an $O(n^3)$ calculation, as each of the n nodes will have $\frac{n(n-1)}{2}$ edges. This is to be balanced with the more favorable Kolmogorov-Smirnov analysis of the normalized entropies NVE, and NVE', which yield lower P-Values. This lower value indicates greater predictive power, but at the expense of a more expensive calculation.

VI. Conclusions

In this paper we introduced computable, node level alternatives to structural entropy measures that are useful when identifying critical nodes in a network. Building on the approach of network science established in Barabási's pivotal paper, and suggestions made in the work of Dehmer, we have advanced computable metrics using structural information available within one hop of a network node. By analyzing

TABLE XI: Kolmogorov-Smirnov Analysis of Null Hypothesis for Event Incident Distributions

Metric	D-stat	D-Crit	α	P Value	Significant
VE	0.5055	0.0396	5%	0.63%	Yes
VE'	0.4624	0.0399	5%	0.26%	Yes
NVE	0.4489	0.0399	5%	0.19%	Yes
NVE'	0.4462	0.0404	5%	0.16%	Yes
CE	0.4665	0.0399	5%	0.29%	Yes
CVE'	0.4394	0.0403	5%	0.14%	Yes
Degree	0.0368	0.0403	5%	9.29%	No

the extremal properties of well known global graph entropies, we were able to identify that they satisfy the criteria required to be a valid entropy, and have similar extremal behavior to the global values when considering special graphs. Critically, the introduction of normalization based upon the clustering coefficient of a nodes neighborhood improves the utility of the metric. We applied these measures to our proprietary data set. Applied to the datasets, we obtain a distribution that isolates a small subset of nodes with high values, a necessary condition to be acceptable as a metric.

This analysis is further supported when we look at the distribution of events and incidents by the value of the metric. We have a clear correlation between high values of the metric and the propensity for the node to produce incidents. This is substantiated by hypothesis testing to eliminate the possibility that the distributions are similar to each other, and therefore that any difference in distribution of events and incidents is purely random. Additional precision and recall analysis using a modified F_β score indicates that there is the possibility of establishing a value of the metric whereby minimal loss of recall (20% of incidents missed) is tolerable to achieve a reduction of 65% in the event rate that needs to be processed. In the context of the large and dynamic networks of current implementations this could be a critical improvement in the performance of root cause algorithms.

All of our analysis has been constrained to the immediate one-hop neighborhood of a node. The justification of studying these values in practical networks has been achieved in theory, and in further work we intend to analyze more real world datasets, and extend our entropy measures to include j-spheres for $j > 1$. In addition, we plan to compare vertex entropy against other node importance measures such as betweenness, to assess the difference in effectiveness as compared to cost of calculation.

REFERENCES

[1] S. Boccaletti, G. Bianconi, R. Criado, C. I. del Genio, J. Gómez-Gardeñes, M. Romance, I. Sendiña-Nadal, Z. Wang, and M. Zanin, "The structure and dynamics of multilayer networks," *Physics Reports*, vol. 544, no. 1, pp. 1–122, 2014. [Online]. Available: http://dx.doi.org/10.1016/j.physrep.2014.07.001

[2] M. L. Steinder and A. S. Sethi, "A survey of fault localization techniques in computer networks," *Science of Computer Programming*, vol. 53, no. 2, pp. 165–194, nov 2004. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0167642304000772

[3] S. Kliger, S. Yemini, and Y. Yemini, "A coding approach to event correlation," ... *Network Management IV*, 1995. [Online]. Available: http://link.springer.com/chapter/10.1007/978-0-387-34890-2_24

[4] M. Miyazawa and K. Nishimura, "Scalable root cause analysis assisted by classified alarm information model based algorithm," ... *of the 7th International Conference on* ..., pp. 2–5, 2011. [Online]. Available: http://dl.acm.org/citation.cfm?id=2147737

[5] R. Harper, "Entropy & The Science of Noisele," 2016. [Online]. Available: https://www.moogsoft.com/whats-new/entropy-noise/

[6] L. Metcalf and J. M. Spring, "Blacklist Ecosystem Analysis Spanning Jan 2012 to Jun 2014," *ACM Digital Library*, pp. 13–22, 2014.

[7] D. Powers, "Evaluation: From Precision, Recall and F-Measure To Roc, Informedness, Markedness & Correlation," *Journal of Machine Learning Technologies*, vol. 2, no. 1, pp. 37–63, 2011. [Online]. Available: http://www.bioinfopublication.org/files/articles/2_1_1_JMLT.pdf

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

[9] M. Faloutsos, P. Faloutsos, and C. Faloutsos, "On Power-Law Relationships of the Internet Topology," *In SIGCOMM*, pp. 251–262, 1999. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.37.234

[10] L. Li, D. Alderson, W. Willinger, and J. Doyle, "A First-Principles Approach to Understanding the Internet ' s Router-level Topology," *Acm Sigcomm*, pp. 3–14, 2004.

[11] B. Bollobás and O. Riordan, "Robustness and Vulnerability of Scale-Free Random Graphs," *Internet Mathematics*, vol. 1, no. 1, pp. 1–35, 2004.

[12] B. Bollobás and O. Riordan, "Mathematical results on scale-free random graphs," in *Handbook of Graphs and Networks*. Wiley-VCH, 2006, ch. Mathematic, p. 417.

[13] R. Albert, H. Jeong, and A. Barabasi, "Error and attack tolerance of complex networks," *Nature*, vol. 406, no. 6794, pp. 378–82, 2000. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/10935628

[14] L. Freeman, "Centrality in social networks conceptual clarification," *Social networks*, vol. 1, no. 1968, pp. 215–239, 1979. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0378873378900217

[15] L. Spizzirri, "Justification and Application of Eigenvector Centrality," *Math.Washington.Edu*, 2011. [Online]. Available: https://www.math.washington.edu/∼morrow/336_11/papers/leo.pdf

[16] U. Brandes, "A faster algorithm for betweenness centrality*," *The Journal of Mathematical Sociology*, vol. 25, no. 2, pp. 163–177, 2001.

[17] P. Tee, G. Parisis, and I. Wakeman, "Towards an approximate graph entropy measure for identifying incidents in network event data," in *NOMS 2016 - 2016 IEEE/IFIP Network Operations and Management Symposium*, April 2016, pp. 1049–1054.

[18] Y. Kanda, R. Fontugne, K. Fukuda, and T. Sugawara, "ADMIRE: Anomaly detection method using entropy-based PCA with three-step sketches," *Computer Communications*, vol. 36, no. 5, pp. 575–588, 2013.

[19] G. Nychis, V. Sekar, D. G. Andersen, H. Kim, and H. Zhang, "An empirical evaluation of entropy-based traffic anomaly detection," *Proceedings of the 8th ACM SIGCOMM conference on Internet measurement conference - IMC '08*, p. 151, 2008. [Online]. Available: http://portal.acm.org/citation.cfm?doid=1452520.1452539

[20] M. E. Bolanos, S. Aviyente, and H. Radha, "Graph entropy rate minimization and the compressibility of undirected binary graphs," *2012 IEEE Statistical Signal Processing Workshop, SSP 2012*, no. 2, pp. 109–112, 2012.

[21] G. Simonyi, "Graph entropy: a survey," *Combinatorial Optimization*, vol. 20, pp. 399–441, 1995.

[22] A. Mowshowitz and V. Mitsou, "Entropy, Orbits, and Spectra of Graphs," *Analysis of Complex Networks: From Biology to Linguistics*, pp. 1–22, 2009.

[23] M. Dehmer and A. Mowshowitz, "A history of graph entropy measures," *Information Sciences*, vol. 181, no. 1, pp. 57–78, 2011.

[24] 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.

[25] P. Tee, I. Wakeman, G. Parisis, and J. Dawes, "Is Preferential Attachment the 2nd Law of Thermodynamics in Disguise?" *ArXiv e-prints*, Dec. 2016. [Online]. Available: https://arxiv.org/abs/1612.03115v2

[26] H. S. Wilf, "The Eigenvalues of a Graph and Its Chromatic Number," *Journal of the London Mathematical Society*, vol. s1-42, no. 1, pp. 330–332, 1967.

[27] H. S. Wilf and G. Szekeres, "An Inequality for the Chromatic Number of a Graph," *Journal of Combinatorial Theory*, vol. 4, no. 1, pp. 1–3, 1968.

[28] J. Körner, "FredmanKomlós bounds and information theory," pp. 560–570, 1986.

[29] 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.
[30] D. Watts and S. Strogatz, "Collective Dynamics of 'Small-World' Networks," *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.
[31] A.-L. Barabási, *Network Science*. Cambridge University Press; 1 edition (August 5, 2016), 2016.
[32] 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.

Philip Tee is the founder, CEO and Chairman of Moogsoft Inc, a pioneering provider of data science enabled Service Management products. He is a serial entrepreneur, having been part of the founding team and principle inventor of the technologies of 4 companies that have gone public or been acquired. Software products he designed continue to manage some of the largest communications networks in the world, and, are in use at over 1000 companies. He is the author of 15 patent applications for methods of fault management. He has a BSc in Chemical Physics and is a PhD candidate in Informatics at the University of Sussex.

Ian Wakeman is a Professor in Software Systems in the Department of Informatics at the University of Sussex. He has a BA in Electrical and Information Sciences from Cambridge University, a MS from Stanford University and a PhD from UCL. His research could be described as user-centred networking, investigating protocols and techniques to make computer networks work for people. This has spawned over 90 refereed papers in fields as diverse as congestion control for packetized video, programming languages for active networks and has more recently focused on communication in challenged environments. He is the co-founder of InCrowd Sports Ltd, which provide app driven connectivity within sports stadia.

George Parisis is a Lecturer in Computer Science at the University of Sussex. He has a BA in Computer Science, MSc in Information Systems and a PhD from the Department of Informatics, Athens University of Economics and Business. His research is in data centre networks and data transport, opportunistic and information-centric networks. He has published over 20 papers in international, peer-reviewed conferences and journals.

Correction to "Vertex Entropy as a Critical Node Measure in Network Monitoring"

Philip Tee
George Parisis and Ian Wakeman

I. Corrected Proof of Proposition 1

From Definition 2 in the paper we have for a simple graph $G = (V, E)$, the following expression for the chromatic entropy $I_C(G)$ of a graph:

$$I_C(G) = \min_{\{C_i\}} -\sum_{i=1}^{N_c} \frac{|C_i|}{n} \log_2 \frac{|C_i|}{n} \tag{1}$$

We seek to prove the following proposition related to this form of entropy.

Proposition 1. *For all connected, simple graphs $G = (V, E)$ of order $n > 3$ the star graph S_n minimizes the chromatic entropy $I_c(G)$.*

Proof. Let G be a connected, simple graph of order n. From the definition of chromatic entropy, we know:

$$I_C(G) = \min_{\{C_i\}} -\sum_{i=1}^{N_c} \frac{|C_i|}{n} \log_2 \frac{|C_i|}{n} \tag{2}$$

$$\text{where } -\sum_{i=1}^{N_c} \frac{|C_i|}{n} \log_2 \frac{|C_i|}{n} \tag{3}$$

$$= -\sum_{i=1}^{N_c} \frac{|C_i|}{n} \log_2 |C_i| + \sum_{i=1}^{N_c} \frac{|C_i|}{n} \log_2(n) \tag{4}$$

$$= -\frac{1}{n} \sum_{i=1}^{N_c} |C_i| \log_2 |C_i| + \log_2 n \tag{5}$$

where N_c is the number of chromatic classes of each coloring under consideration. In order to minimize the chromatic entropy, the expression

$$F_{N_c} = \sum_{i=1}^{N_c} |C_i| \log |C_i| \tag{6}$$

must be maximized for all possible colorings and graphs (we have used the fact that $\log_2(x) = \log(x)/\log(2)$ so that $\log_2$ is maximized whenever $\log$ is maximized; we will hence use the natural $\log$ from now on).

Because G is connected it follows immediately that any coloring of G needs to have at least 2 chromatic classes. In the case of 2 chromatic classes, maximizing F_2 reduces to finding the maximum of

$$c_1 \log c_1 + (n - c_1) \log(n - c_1) \tag{7}$$

with respect to c_1 and with the boundary conditions $1 \leq c_1 \leq n-1$. Taking the derivative with respect to c_1 we find the well-known result that the only extremum is at $c_1 = \frac{n}{2}$ and that this is a minimum (as confirmed by the second derivative). Hence, the maximum is achieved at the two (equivalent) boundaries $c_1 = 1$ and $c_1 = n - 1$. The star graph S_n achieves this maximum of $F_2 = (n-1)\log(n-1)$.
It remains to be shown that no other graph with some coloring of 3 or more chromatic classes might achieve a higher value for F_{N_c}. We show this by induction.

Lemma 1. *The maximal value for F_{N_c} for any coloring with $N_c \geq 2$ chromatic classes is achieved when $c_i = |C_i| = 1$ for all but one class and $c_j = |C_j| = n - (N_c - 1)$. The value of F_{N_c} is then $(n - (N_c - 1)) \log(n - (N_c - 1))$.*

Proof. We prove the lemma by induction.
Induction start: Let $N_c = 2$. This is the case discussed above.
Induction assumption: The claim is true for N_c chromatic classes.
Induction step: For $N_c + 1$ chromatic classes, we consider a split where the first class has size c_1, and the remaining classes are thus of such size that $\sum_{i=2}^{N_c+1} c_i = n - c_1$. We can split the expression for F_{N_c+1} into

$$F_{N_c+1} = c_1 \log c_1 + \sum_{i=2}^{N_c+1} c_i \log c_i \tag{8}$$

We already know that the second term is maximized, for any given value of c_1, when the remaining N_c classes split into $N_c - 1$ classes of size 1 and one class of size $n - (N_c - 1) - c_1$ and hence the maximal value for F_{N_c+1} as a function of c_1 is:

$$F_{N_c+1} = c_1 \log c_1 + (n - (N_c - 1) - c_1) \log(n - (N_c - 1) - c_1) \tag{9}$$

$$= c_1 \log c_1 + (n' - c_1) \log(n' - c_1) \tag{10}$$

where we have denoted $n - (N_c - 1) = n'$. The resulting expression as a function of c_1 is the same as equation (7) investigated above and its maximum with respect to c_1 is hence achieved for $c_1 = 1$ or $c_1 = n' - 1$. Both of these solutions lead to the final solution claimed and to $F_{N_c+1} = (n - N_c) \log(n - N_c)$ as required. □

The lemma shows that the maximum value for F_{N_c} for any number N_c of chromatic classes is $(n-(N_c-1)) \log(n-(N_c-1))$ which is decreasing in N_c and hence has its maximal value for the minimal possible value $N_c = 2$. This demonstrates that no other graph can possibly do better than the star graph S_n which achieves $N_c = 2$ and $c_1 = 1$ on its best coloring, leading to the maximal value of F_{N_c} and hence minimal chromatic entropy. □

ACKNOWLEDGMENT

The authors would like to thank Professor Thomas Nowotny who supplied the corrected proof.

5.2 Discussion

5.2.1 Comparing Vertex Entropy and Graph Entropy of Sampled Graphs

In the paper [75], and the paper presented in this chapter, it is referenced that the various proposed vertex entropy measures behave in analogous ways to known global measures of graph entropy when summed across a whole graph. In both papers the analytical comparison is limited to the exploration of the extremal behavior of the vertex measures as compared to the extremal behaviors of the global metrics. No attempt is made to state or prove that the local measures are an accurate approximation to global entropy, however the question of how closely they approximate global measures was left as an open question. One way to establish the correlation is to numerically compare the vertex and global measures against a range of randomly generated graphs. We present in this section the results of such an analysis.

Experimental Approach

The calculation of the chromatic information content and structural entropy is known to be *NP*-complete. We can make use of the following relation from [50] for simple, connected, undirected graphs to reduce the computation load to just the calculation of Chromatic Information::

$$H(G) = \log_2(n) - I_C(G) \tag{5.1}$$

To compute $I_C(G)$ we can exploit a greedy algorithm, as described in [47] to calculate the the approximate chromatic classes of a graph. The value obtained by the greedy algorithm will produce a value for the chromatic number, and hence chromatic classes, which will be bounded below by the actual value for the graph. Nevertheless it is computable for relatively large graphs and the value of chromatic information obtained is a reasonable approximation to the actual value for the graph. Calculating the optimal chromatic coloring of a graph is not computable with the exception of trivial graphs of small values of n. Once the chromatic entropy, $I_C(G)$, is computed it is simple to calculate structural entropy using equation (5.1).

In our analysis, we consider two classes of random graphs, the classic Erdős-Rényi random graph $G(N, p)$ on N nodes with link probability p , and scale free graphs generated with preferential attachment following the scale free model of Barabási Albert, as described in [14, 6]. Simulating using the two different approaches is considered valuable as scale free graphs are considered more realistic models of real networks, and traditional random graphs will have very different clustering properties, which will produce different types of

graph colorings. For the case of Erdős-Rényi random graphs we produce a sample of graphs with a varying value of p, to produce graphs with increasing density of edges as p increases, but with a fixed number of 100 vertices. For the scale free graphs, we produce two separate samples. In the first sample of graphs with 300 nodes, we vary the new node connections parameter m which results in graphs of increasing edge density, and in the second sample we vary the number of vertices from 70 to 270, and adjust the value of m to produce graphs with approximately similar edge density. This approach is taken to isolate whether varying edge density or graph size breaks the correlation between vertex entropy measures and global graph entropy.

Generating the graphs for analysis was performed using the simulator program `pa_simulator` developed during the research program in two different modes. For the Erdős-Rényi graphs a series of 100 node graphs were generated varying the connection probability between 0.3 and 0.7, incrementing by 0.2. The value of 0.3 was chosen as this places the random graph in the super critical regime of the graph where the giant component of the graph is likely to contain all 100 nodes (as described in [6]), and at greater than 0.7 the graphs become progressively completely connected. This experiment specifically tests the effect of edge density on the correlation.

For the scale free graphs we sampled a number of 300 node graphs, stepping the attachment parameter m from 2 to 23. This generated graphs with edge densities that range from 2% to 25% fully connected (a fully connected 300 node graph has $44,850$ edges). This choice of range was chosen to produce a giant component containing over 90% of the nodes in the graph, whilst not producing a graph with too high an edge density. Additionally we produced a sample of graphs of increasing size (70 to 270 nodes, whilst adjusting the parameter m to maintain a constant edge density of approximately 94%. This experiment specifically isolates the effect of vertex count on any correlation between vertex and global entropy measures, across a selection of similarly densely connected graphs.

Analysis

In Figures 5.1, 5.2 and 5.3 we present plots of both Structural and Chromatic entropy against sums of vertex entropy for the principal four measures studied (VE(v), VE'(v), NVE(v), NVE'(v)). Analysis was not conducted on the cluster entropy measures described in the paper, which were introduced as comparative variants of the first four versions of vertex entropy. In every case, there is evidence of a correlation between the whole graph entropy measures and the vertex entropy measures when summed across the whole graph. This approximate correlation is intriguing, and may indicate a relationship between the global

graph entropy measures and the local vertex approximations. Whilst falling short of a proof of an exact relationship between the two approaches to graph entropy, the result motivates further investigation and certainly supports and substantiates the use of the vertex entropy measure in the papers presented in Chapters 4 and 5.

In further work it would be interesting to investigate this relationship with additional simulations, and for other measures of both vertex and global entropy. In particular no account is taken of Von Neumann entropy as a global measure (as described in [57]), and of course vertex entropy described on larger j-Spheres. It is also possible that other node importance measures such as centrality may also correlate well with both global and vertex measures of entropy.

5.3 F_β Analysis and ROC curves

5.3.1 F_β Analysis with β=100

Standard F_1 score analysis [59], uses the harmonic mean of precision and recall to assess the effectiveness of an algorithm as a binary classifier. It can be written as:

$$F_1 = \frac{2}{\left(\frac{1}{precision} + \frac{1}{recall}\right)}, \tag{5.2}$$

and consequently the value of the F_1 score ranges from $0 < F_1 \leq 1$, with 1 being the case of perfect recall and precision. Implicit in this definition is an equal weighting of both precision and recall, so there is no way to distinguish from a sub-optimal F_1 score whether it is recall or precision that has caused the value to be less than 1. In addition the samples are extremely unbalanced as we are comparing approximately $1,500$ incidents against $23,000,000$ events, where an incident is counted as a positive outcome, and an event could be either positive or negative. In many applications one of the two measures is more practically important, and to capture that the F_β extension introduces a free parameter β to allow the score to be biased accordingly. The definition as given in Equation (1) of the paper presented in this chapter is:

$$F_\beta = (1+\beta^2) \times \frac{precision \times recall}{\beta^2 \times precision + recall}, \tag{5.3}$$

which for $\beta > 1$ biases recall and for $\beta < 1$ biases precision.

In the case of our application to events and incidents, the cost of a missed incident is far higher to a network operator than the time spent analyzing a captured incident that

contains spurious events. Typically the ratio of events to incidents is extremely large (in the data set analyzed over 96 million events that produced approximately $37,000$ incidents, or nearly $2,600$ events per incident. Although a typical value of β in many applications is 2, to accurately capture this wide disparity between the number of events and incidents, it was felt a much higher value of β was required. The choice of $\beta = 100$, was motivated by the fact that in most commercial fault management applications a $100:1$ compression of events to alerts is expected by the simple application of de-duplication techniques [36, 25, 48]. As $\beta \to \infty$, Equation (5.3) collapses to equate to the value of recall, and a concern of choosing $\beta = 100$ is that the F_β score obtained is not sufficiently distinguished from recall to provide any further insight. In Figure 5.4, we present a plot of the F_β scores for $\beta = 10$ and $\beta = 100$, alongside recall for one of our metrics, normalized fractional vertex entropy, $NVE'(v)$. As expected, at high values of vertex entropy the measures all converge, but for low values of the metric there is a considerable difference between recall and F_β, including the absence of a maximum. It is this maximum that is practically useful for choosing an optimal vertex entropy threshold, and I conclude that the F_β analysis is a useful approach.

5.3.2 ROC Curve Analysis

In the paper we present the analysis of the effectiveness of the vertex entropy cutoff using the F_β analysis described above. The experimental data, is however, amenable to the use of a Receiver Operating Characteristic (ROC) curve approach, (an excellent survey of the ROC approach is presented in Powers *et al* [59]). We have performed this analysis and present the results here. Unfortunately the paper was too far progressed in the publication process to substitute wholesale the F_β analysis and length constraints did not permit the inclusion.

It is important to remember that the vertex entropy metrics are not intended to be used as a standalone binary classifier, but as part of a pipeline of algorithms that effectively classify events as causal or non-causal. Instead the algorithms are intended to pre-process the event data to remove load from the downstream fault localization techniques that we surveyed in Chapter 1. ROC curves plot the value of the False Positive Rate (FPR) against the True Positive Rate (TPR), which we obtain from the incident and event data. In this data, incidents are assumed to indicate the presence of a causal event, that we deem a positive outcome for the classifier. However, the incidents and events are only linked by their node attributes, for which we have a value of each entropy metric. We do not know which events caused the incident, but the presence of an incident means that at least one of the events from that

node is causal. In our analysis we assume that one event from an incident producing node is causal.

The entropy metrics are proposed to be used as a filter threshold, such that any events originating from nodes with an entropy value below the threshold are discarded. The purpose of the ROC curve is therefore to measure the effectiveness of this threshold at classifying events as either causal or non-causal.

The experimental points in the ROC curve that we will generate, correspond to choices of the given vertex entropy, and, before embarking on the analysis it is helpful to define the terms and relate them back to the data we are analyzing, which we do in Table 5.1 below.

Metric	Definition	Method
True Positive (TP)	Number of correct positive predictions of classifier	The number of events that are mapped to respective incidents from nodes above or at the cutoff point. Based upon the assumption of one causal event per incident, this is numerically equal to the number of incidents at or above the threshold.
True Negative (TN)	Number of correct negative predictions of classifier	The number of events from nodes below the cutoff point excluding the ones for which there is a recorded incident.
False Positive (FP)	Number of incorrect positive predictions of classifier	The number of events from nodes above or at the cutoff point excluding the ones for which there is a recorded incident.
False Negative (FN)	Number of incorrect negative predictions of classifier	The number of events that are mapped to respective incidents from nodes below the cutoff point. Based on the assumption of one causal event per incident, this is numerically equal to the number of incidents from nodes below the cutoff point.

Table 5.1 Calculation of ROC Components in Event and Incident Data

Using these definitions we can make a reliable estimate of the True Negative Rate (TNR) and make use of the identity $FPR = 1 - TNR$ from [16]. The calculation for TNR, for a value of entropy $h \in [0, h_{max}]$ is:

$$TNR = \frac{TN(h)}{TN(h) + FP(h)} \,. \tag{5.4}$$

Using this approach, the data was linearly bucketed into $1,000$ buckets of events and incidents for equal increments of each entropy metric concerned, from zero to the maximum value. Unfortunately most of the metrics exhibit a very significant cluster of events at the zero value of the metric, which is not amenable to adjustments to the bucketing strategy, to for example a logarithmic, or fixed event/incident count buckets. This inevitably means that there is a gap in the ROC curves between the highest false positive rate and the $(1,1)$ point of maximum FPR and TPR.

In addition to plotting the ROC curves, numerical trapezoidal integration was used to calculate the Area Under the Curve (AUC) of the ROC plots. This is a well established measure of the effectiveness of a binary classifier, and is linked to the probability of a randomly chosen positive example being correctly labelled, as compared to a randomly chosen negative example. The values are presented in Table 5.2.

Metric	AUC
VE	0.7619
VE’	0.7204
NVE	0.7863
NVE’	0.7824
CE	0.7296
CVE’	0.7072

Table 5.2 Area Under the ROC Curve (AUC) Analysis of Entropy Measures

In Figure 5.5 the ROC curve plots for each of the vertex entropy types are presented. Despite the fact that the vertex entropy measures are not intended to be used as a standalone binary classifier, the measures perform reasonably well. For example, using the AUC measure, a perfect classifier would have an AUC of 1.0, which would indicate that with 0% false positives 100% recall is achieved. As a heuristic any classifier with > 80% AUC would be considered excellent or good (see for example [77], [16]), and values of less than 70% being considered poor and 50% effectively a random classifier. In the case of our metrics, the normalized entropy measures NVE and NVE' come close to 80% and may even have utility

as a standalone classifier on that basis. In any case this analysis provides further evidence that the metrics are effective as a noise reduction step, prior to a causal analysis algorithm such as those described in Chapter 1.

5.3.3 Corrected Proof of Proposition 1

Post the publication of the paper, discussions with colleagues highlighted a flaw in the proof of Proposition 1. The proof rests upon all graphs of order n being achievable from either a Star Graph S_n or a Path, P_n. This is not the case and can be demonstrated by considering the following graph of order 5, which cannot be created by adding edges to either S_5, or P_5:

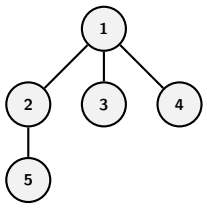


Fig. 5.6 A Counter Example Graph to the Claim in Proposition 1.

However, suggestions from colleagues provided a correct proof. Unfortunately the paper was already accepted and in post production, so I was not able to amend the proof in the original paper, but I have submitted an erratum to the journal which is being processed. I outline the new proof here:

Proposition 1. *For all connected, simple graphs $G = (V,E)$ of order $n > 3$ the star graph S_n minimizes the chromatic entropy $I_c(G)$.*

Proof. Let G be a connected, simple graph of order n. From the definition of chromatic entropy, we know:

$$I_C(G) = \min_{\{C_i\}} -\sum_{i=1}^{N_c} \frac{|C_i|}{n} \log_2 \frac{|C_i|}{n} \tag{5.5}$$

$$\text{where } -\sum_{i=1}^{N_c} \frac{|C_i|}{n} \log_2 \frac{|C_i|}{n} \tag{5.6}$$

$$= -\sum_{i=1}^{N_c} \frac{|C_i|}{n} \log_2 |C_i| + \sum_{i=1}^{N_c} \frac{|C_i|}{n} \log_2(n) \tag{5.7}$$

$$= -\frac{1}{n} \sum_{i=1}^{N_c} |C_i| \log_2 |C_i| + \log_2 n \tag{5.8}$$

where N_c is the number of chromatic classes of each coloring under consideration. In order to minimize the chromatic entropy, the expression

$$F_{N_c} = \sum_{i=1}^{N_c} |C_i| \log |C_i| \tag{5.9}$$

must be maximized for all possible colorings and graphs (we have used the fact that $\log_2(x) = \log(x)/\log(2)$ so that $\log_2$ is maximized whenever log is maximized; we will hence use the natural log from now on).

Because G is connected it follows immediately that any coloring of G needs to have at least 2 chromatic classes. In the case of 2 chromatic classes, maximizing F_2 reduces to finding the maximum of

$$c_1 \log c_1 + (n - c_1) \log(n - c_1) \tag{5.10}$$

with respect to c_1 and with the boundary conditions $1 \leq c_1 \leq n-1$. Taking the derivative with respect to c_1 we find the well-known result that the only extremum is at $c_1 = \frac{n}{2}$ and that this is a minimum (as confirmed by the second derivative). Hence, the maximum is achieved at the two (equivalent) boundaries $c_1 = 1$ and $c_1 = n-1$. The star graph S_n achieves this maximum of $F_2 = (n-1)\log(n-1)$.
It remains to be shown that no other graph with some coloring of 3 or more chromatic classes might achieve a higher value for F_{N_c}. We show this by induction.

Lemma 1. *The maximal value for F_{N_c} for any coloring with $N_c \geq 2$ chromatic classes is achieved when $c_i = |C_i| = 1$ for all but one class and $c_j = |C_j| = n - (N_c - 1)$. The value of F_{N_c} is then $(n - (N_c - 1)) \log(n - (N_c - 1))$.*

Proof. We prove the lemma by induction.
Induction start: Let $N_c = 2$. This is the case discussed above.
Induction assumption: The claim is true for N_c chromatic classes.
Induction step: For $N_c + 1$ chromatic classes, we consider a split where the first class has size c_1, and the remaining classes are thus of such size that $\sum_{i=2}^{N_c+1} c_i = n - c_1$. We can split the expression for F_{N_c+1} into

$$F_{N_c+1} = c_1 \log c_1 + \sum_{i=2}^{N_c+1} c_i \log c_i \tag{5.11}$$

We already know that the second term is maximized, for any given value of c_1, when the remaining N_c classes split into $N_c - 1$ classes of size 1 and one class of size $n-(N_c-1)-c_1$ and hence the maximal value for F_{N_c+1} as a function of c_1 is:

$$F_{N_c+1} = c_1 \log c_1 + (n-(N_c-1)-c_1)\log(n-(N_c-1)-c_1) \tag{5.12}$$
$$= c_1 \log c_1 + (n'-c_1)\log(n'-c_1) \tag{5.13}$$

where we have denoted $n-(N_c-1) = n'$. The resulting expression as a function of c_1 is the same as equation (5.10) investigated above and its maximum with respect to c_1 is hence achieved for $c_1 = 1$ or $c_1 = n'-1$. Both of these solutions lead to the final solution claimed and to $F_{N_c+1} = (n-N_c)\log(n-N_c)$ as required. □

The lemma shows that the maximum value for F_{N_c} for any number N_c of chromatic classes is $(n-(N_c-1))\log(n-(N_c-1))$ which is decreasing in N_c and hence has its maximal value for the minimal possible value $N_c = 2$. This demonstrates that no other graph can possibly do better than the star graph S_n which achieves $N_c = 2$ and $c_1 = 1$ on its best coloring, leading to the maximal value of F_{N_c} and hence minimal chromatic entropy. □

5.3.4 Normalization of Inverse Degree Entropy

In section IV.A of the paper there is a discussion regarding the use of $p_i = \frac{1}{k_i}$ as a probability functional, and in equation (17) I present an expression for the value of its sum over all vertices. The probability definition of in Dehmer's original treatment [23] would more clearly have been represented, as suggested post publication, by writing equation (14) as:

$$p_i = \frac{\frac{1}{k_i}}{\sum_i^n \frac{1}{k_i}} \tag{5.14}$$

The denominator of this expression sums to a constant value for all vertices, C. If we write $p_i = \frac{1}{k_i}$, this will break the normalization of the sum over all vertex probabilities, and so needs to be divided by C to be a valid probability. The effect of including the normalization constant would complicate the subsequent discussion of the inverse degree entropy, and so I took the decision to exclude it. The motivation was that the essential aspects of the complexity of the graph structure is captured by just using the inverse degree of the vertex as the probability in the entropy measure. In retrospect this reasoning could have been more explicit in the paper, but was reduced in rigor due to length constraints.

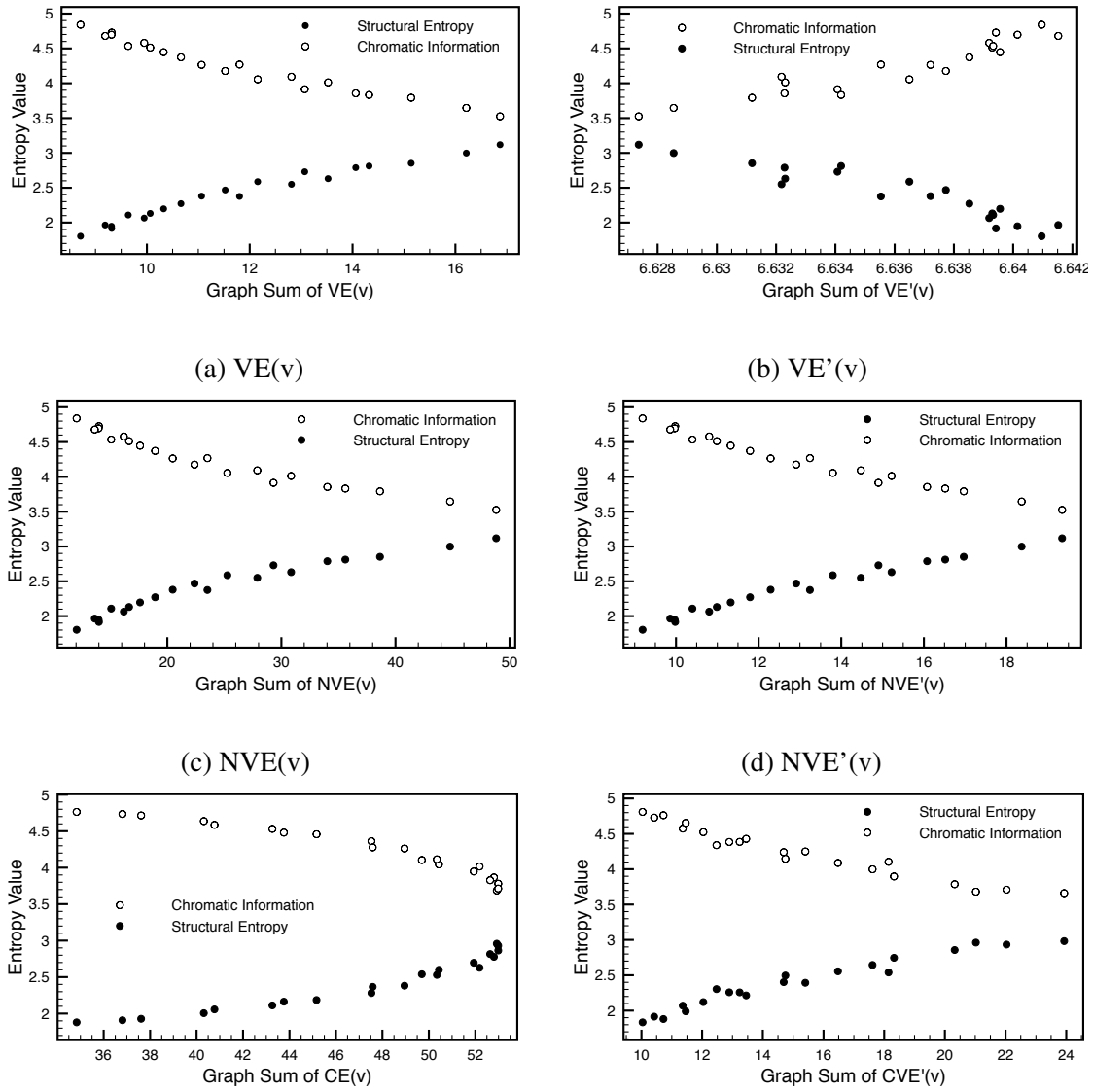
5.3.5 Shannon's Desiderata

In standard texts on entropy such as 'Information theory' by J.Stone [69] and the work of Csiszár [22], a number of constraints are enumerated upon a valid entropy metric. These constraints, or 'desiderata', are used to argue the necessary form of the entropy equation, and in particular the need for the introduction of logarithms to guarantee additivity, symmetry, maximality and positivity. In the Definition 1 of the paper, I reproduce these requirements as a test to ensure that any entropy metrics subsequently defined are well behaved relative to the 'desiderata'.

In the case of symmetry, the metric automatically inherits the intrinsic symmetry of the graph union operation, and therefore the requirement is always trivially satisfied. The inclusion was made to underline the analogy in reasoning between graph entropy and standard Shannon entropy. The maximality condition could also have been more unambiguous in regard to the uniqueness of the maxima. Uniqueness is not proven in the text of the paper, nor is it essential to the scientific content of the contribution. This point could have been made more concretely in the definition. Additionally, as the graphs considered are finite, the existence of a maximum is trivially satisfied.

5.3.6 Definition of the j-Sphere

Definition 4 of the paper contains a potentially confusing, and unnecessary union with the central node v_i. As the distance condition is written as $d(v_i, v_j) \leq 1$, the central vertex is automatically captured. Originally written to stress the inclusion of the vertex, in retrospect this can obscure the definition. Additionally the condition $j \geq 1$ inside the set definition is not good practice in defining sets, despite being the notation employed by Dehmer in his series of papers [23, 24]. Definition 4 should be more correctly be written as it is in Chapter 2, Definition 8.



(a) VE(v)

(b) VE'(v)

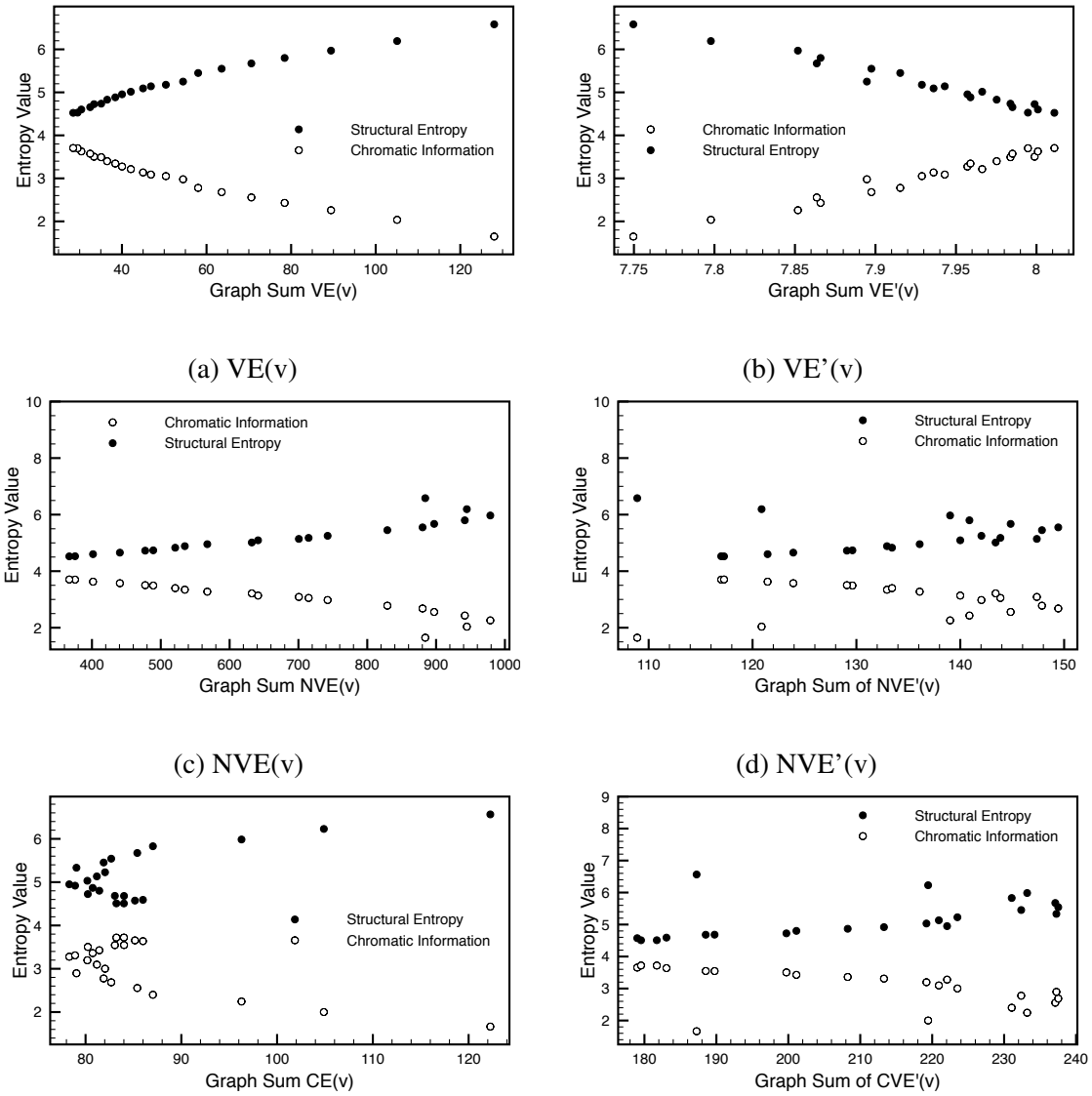
(c) NVE(v)

(d) NVE'(v)

(e) CE(v)

(f) CVE'(v)

Fig. 5.1 Sampled sum of Vertex Entropies for $G(N,p)$ Erdős-Rényi Graphs, with $p \in [0.3, 0.7]$ and $|V| = 100$



(a) VE(v)

(b) VE'(v)

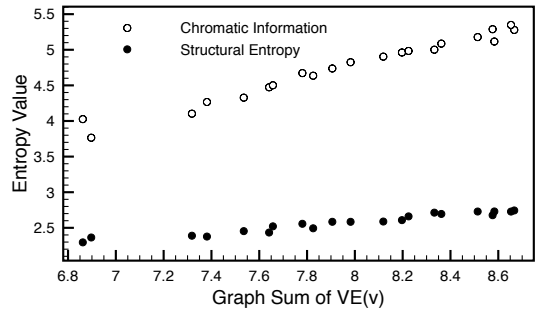
(c) NVE(v)

(d) NVE'(v)

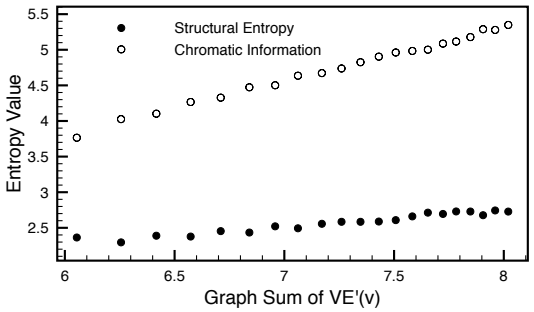
(e) CE(v)

(f) CVE'(v)

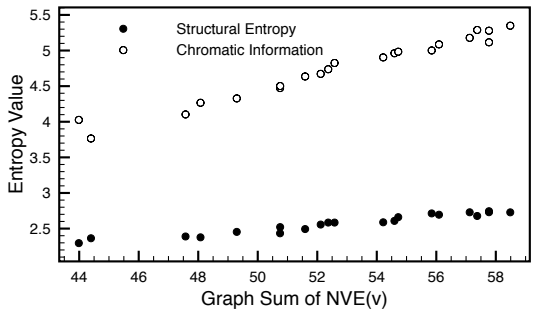
Fig. 5.2 Sampled sum of Vertex Entropies for Scale Free Graphs with $m \in [2,23]$ and $|V| = 300$



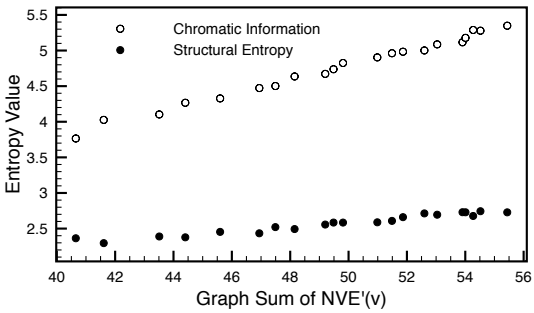
(a) VE(v)



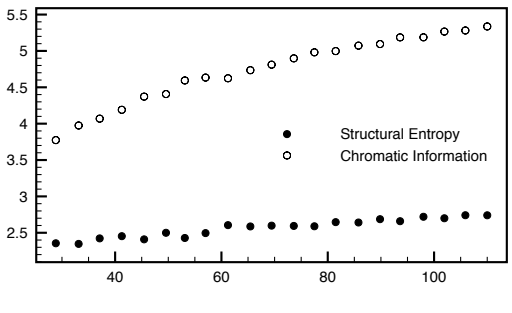
(b) VE'(v)



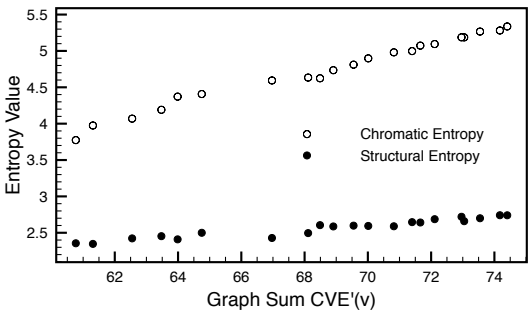
(c) NVE(v)



(d) NVE'(v)



(e) CE(v)



(f) CVE'(v)

Fig. 5.3 Sampled sum of Vertex Entropies for Scale Free Graphs with for $|V| \in [70, 270]$, and edge density 94-98%

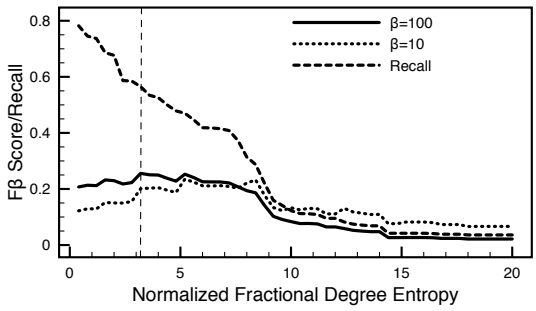
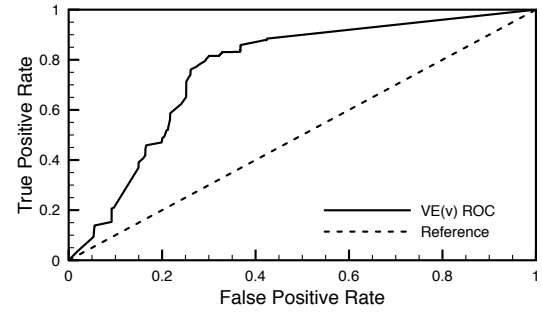
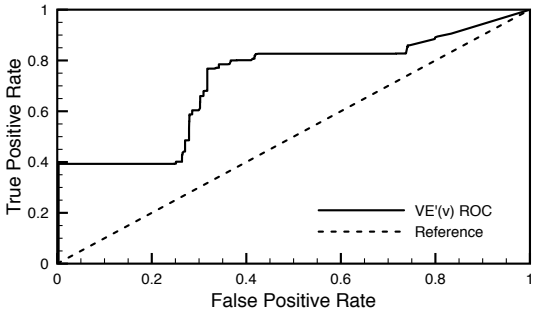


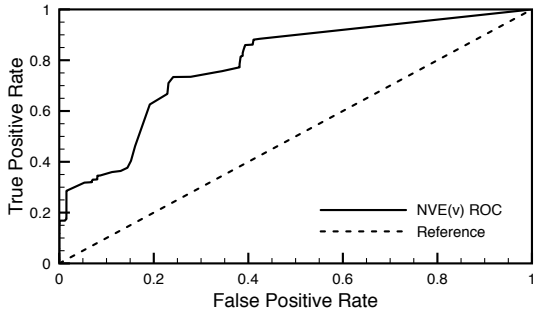
Fig. 5.4 NVE'(v) F_β Plots against Recall



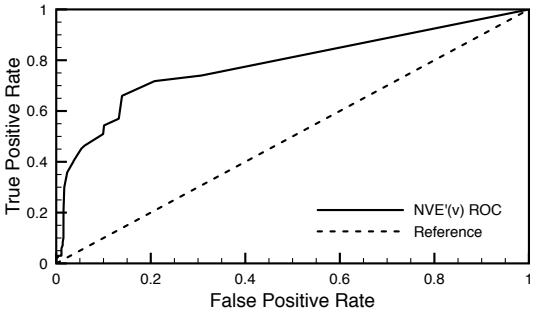
(a) ROC curve for VE(v)



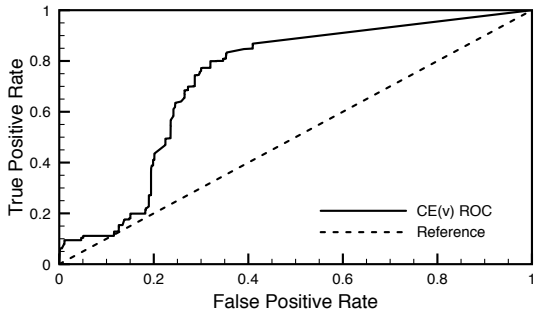
(b) ROC curve for VE'(v)



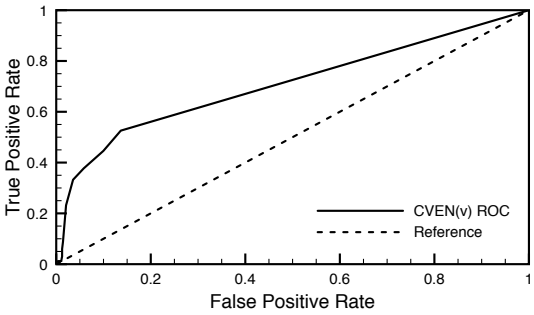
(c) ROC curve for NVE(v)



(d) ROC curve for NVE'(v)



(e) ROC curve for CE(v)



(f) ROC curve for CVE'(v)

Fig. 5.5 ROC Curve Analysis for Event and Incident Distribution for each Vertex Entropy Metric

Chapter 6

Constraints and Entropy in a Model of Network Evolution

6.1 Background to Third Publication

6.1.1 Motivation and Summary of Contribution

Towards the end of the research into the relationship between vertex entropy and incident production, I became interested in the departure of network structure from canonical models such as preferential attachment [2]. In particular, at the MoN15 conference it was clear, that although the preferential attachment model was regarded as effective at explaining the macroscopic behavior of dynamic network growth, it has many shortcomings when applied to actual network data (there are many references but [9], [28] outline the main arguments, and indeed Section VIII of [2] summarizes many of the previous avenues of attack).

A common feature of communications networks is the presence of a physical constraint. These constraints are often practical, or economic in nature, and manifest as a limitation in the number of connected devices a network switch can support. This simple fact is in opposition to a fundamental principle of the preferential attachment that a node can indefinitely acquire new connections. This motivated the exploration of an alternative, *constrained attachment*, model in which the presence of a hard constraint could be elevated to an axiom of the model.

After developing the model, and observing that the results were in good agreement with the initial data available, I expanded the analysis to include many other types of networks. A primary source of such data was found in the Stanford Large Network Dataset [43], which contained a broad collection of real network data. The agreement between the new model and experimental data remained excellent.

Despite the empirical support for the constrained attachment model, I still felt that the approach of postulating an attraction mechanism to be unsatisfactory. It would be more elegant if an underlying physical reason for nodes to preferentially seek attachment to other nodes could be found. Motivated by the work of [56], I began to explore the relationship of vertex entropy to preferential attachment, and in Section 4 of this paper I presented the culmination of my research efforts, which demonstrated that indeed preferential attachment could be derived as a consequence of the universal law of maximal entropy.

An initial presentation of the work prompted discussions with Jonathan Dawes and István Kiss, which led to their collaboration in the production of this paper, and I outline their contribution in Section 6.1.4.

6.1.2 Theoretical Contribution

The paper presented describes two novel contributions to the understanding of dynamic network evolution:

- *Constrained Model*: I proposed a new form of the preferential attachment probability that incorporates the notion of a maximum node degree. This is motivated by the considerations outlined above. The model is a simple extension to the canonical form of preferential attachment, but benefits from having only one additional parameter - the maximum node degree. Comparison of the predictions of this model against both simulated and real data confirms both its consistency and validity as a model of real networks. Crucially it establishes the concept that the correct form of the preferential attachment probability may contain higher powers of node degree.

- *Entropic Model*: To conclude the paper I describe a potential mechanism describing how preferential attachment could arise as a mechanism to maximize the entropy of the graph. From our first and second papers we have an expression for the entropy of an individual node of a graph, and node that in one particular case it is maximized as graphs tend towards highly non-uniform, i.e. disordered, graphs. We can then use this expression to derive a probability of attachment, which we note for the early evolution of the degree of a node is approximately similar to the canonical preferential attachment expression.

The two models in the paper neatly unify the work on dynamic network evolution and vertex entropy.

6.1.3 Data and Methods Used

The data analyzed was derived from a number of sources, building upon the data presented in the first two papers. In addition to the digital service provider topology, and ITZ, we also analyzed:

- *The Stanford Large Network Dataset* [43],[45],[46],[44], and [70]: The Stanford Network Analysis Project (SNAP), maintains a large repository of network topology covering a wide range of networks. These include citation graphs, social gaming user graphs, road networks, website linkage graphs and physical network topologies.
- *The Twitter Follower Network* [31]: This is a very large dataset comprising a snapshot of the Twitter follower network and comprises 1.9 billion edges. Each edge represents a followership relation between the nodes, and the nodes represent a unique user.
- *The Openflights Network* [58]: This dataset is a maintained repository of all commercial flight routes between airports across the world. This builds a graph where each node is an airport and each edge an available flight. The analyzed dataset comprised 61,000 routes.
- *The Published Data by Barabási et al* [2]: The publications of Barabási and Albert, summarize many useful statistics including average degree, power law exponent and max degree for a wide variety of networks. Although not directly analyzable in the same way as the primary datasets, it is useful as a source of parameters for simulation and comparison with the theoretical model.

The data is generally available as a CSV file of edges, and was analyzed using the same tools described in Section 4.1.3. A key addition was made to create a graph simulator, `pa_simulator` that creates graphs by simulating the operation of the constrained attachment model. This can be parameterized by supplying values for the maximum node degree (c), the starting size of the seed graph (m_0), and the number of nodes to attach to at each time-step (m). The simulator begins with a fully connected graph of size m_0, and then adds a node at each iteration step that connects with a probability proportional to Equation (12) of the paper. The output of the simulator is the degree distribution of the simulated graph, used to produce the results in Figures 1, 3a, 3b, 4a, 4b, 5a and 5b of the paper.

The primary objective of the experimental work was to obtain degree distributions from each of the datasets and measure the power law exponent γ. Measuring the power law exponent was done following the methodology outlined in Clauset *et al* [19], produced the results in Table 2 of the paper.

6.1.4 Contributions from Co-Authors

The paper benefited from many helpful comments from all of the co-authors, and of course my supervisors provided the early feedback on the model and approach. István Kiss helped me refine the exposition of the constrained attachment model, particularly after the initial comments from the reviewers spotted an inconsistency in the original form of the constrained attachment probability. He also provided useful advice in the construction of the network simulator and use of the results to demonstrate the validity of the model. Jonathan Dawes provided the analytical solution to Equation (24) of the paper and the subsequent simulations and results presented in Figure 6 of the paper. Jonathan also contributed to the general refinement of the arguments in the entropic model and overall production of the paper.

6.1.5 Related Work

The study of dynamic network evolution is a very richly researched topic. Since the publication of the initial work on preferential attachment, there have been many attempts to challenge and extend the model. Much of the related work is summarized beautifully in [2] and in Barabási's recent text [6], but it is worth pointing to two models in particular. Firstly, introduced independently by Dorogovtsev [27] and Krapivsky [42] were models that explicitly incorporated a value of γ less than 3, which is the classic result of the preferential attachment model. This is accomplished using a variable initial attractiveness of nodes which is not strictly dependent on the degree of the node, and in the case of Krapivsky by specifying the value of γ.

The second model is the fitness model of Bianconi described in [7] and [11]. This model allows each node to have a fitness parameter that is used to scale the attractiveness of the node in the attachment probability. The model is not analytically soluble in the same way as preferential attachment, but can be approximately solved if assumptions are made about the distribution of attractiveness across the network. In those cases values of γ less than three and exponential cut-offs are all achievable.

A key part of the argument in the paper is that the constrained model reproduces a better fit to real world data without having to introduce artificial parameters such as described in the models above. It is also derived directly from considerations of network design. This more fundamental approach to the problem of network evolution in the entropic model also gives good agreement with constrained attachment model and could be the basis of a unifying framework for dynamic network evolution.

EPJ manuscript No.
(will be inserted by the editor)

Constraints and Entropy in a Model of Network Evolution

Philip Tee[1,2], Ian Wakeman[2], George Parisis[2], Jonathan Dawes[3], and István Z. Kiss[4]

[1] Moogsoft Inc,
1265 Battery Street, San Francisco, California 94111 USA e-mail: phil@moogsoft.com
[2] School of Engineering and Informatics,
University of Sussex, Brighton BN1 9RH UK e-mail: ianw,g.parisis,p.tee@sussex.ac.uk
[3] Centre for Networks and Collective Behaviour, and
Department of Mathematical Sciences,
University of Bath, Bath BA2 7AY, UK e-mail: j.h.p.dawes@bath.ac.uk
[4] Department of Mathematics, School of Mathematical and Physical Sciences,
University of Sussex, Brighton BN1 9QH, UK e-mail: i.z.kiss@sussex.ac.uk

the date of receipt and acceptance should be inserted later

Abstract. Barabási-Albert's 'Scale Free' model is the starting point for much of the accepted theory of the evolution of real world communication networks. Careful comparison of the theory with a wide range of real world networks, however, indicates that the model is in some cases, only a rough approximation to the dynamical evolution of real networks. In particular, the exponent γ of the power law distribution of degree is predicted by the model to be exactly 3, whereas in a number of real world networks it has values between 1.2 and 2.9. In addition, the degree distributions of real networks exhibit cut offs at high node degree, which indicates the existence of maximal node degrees for these networks.
In this paper we propose a simple extension to the 'Scale Free' model, which offers better agreement with the experimental data. This improvement is satisfying, but the model still does not explain *why* the attachment probabilities should favor high degree nodes, or indeed how constraints arrive in non-physical networks. Using recent advances in the analysis of the entropy of graphs at the node level we propose a first principles derivation for the 'Scale Free' and 'constraints' model from thermodynamic principles, and demonstrate that both preferential attachment and constraints could arise as a natural consequence of the second law of thermodynamics.

1 Introduction and Background

1.1 Overview

The *'Scale Free'* model of Barabási-Albert [1] is widely accepted as the definitive model of how real world networks evolve. This and other dynamic network models consider real world networks as graphs $G(V, E)$, where $V(t)$ is the set of vertices and $E(t)$ the set of edges. Its success at overcoming the difficulties of applying the Erdős-Rényi (ER) random graph model (for a detailed description see [2]) to real world networks is well understood. In particular the model naturally results in a power law degree distribution, as opposed to the random graph model, which has a binomial distribution of node degree, which in the continuum limit of a very large network is approximately Poisson, with well defined higher statistical moments that establish the 'scale' of the graph. This is in stark contrast to the scale free model which does not have well defined moments above the mean. The model described by Barabási-Albert in [3] and [1] builds upon, and provides an explanation for, the notion of the small world network, first introduced by Watts and Strogatz [4] and has been used to analyze a wide variety of real world graphs.

On close examination, the scale free model has a number of theoretical challenges, and, it is well understood that the behavior of real world networks has deeper complexity than a single constant power law degree distribution. Of course balanced against the success of the model in generating networks that share the *small world* property and scale free degree distributions, these challenges can be viewed as opportunities for refinement of the fundamental approach. In this work we focus on extensions to the model which provide improvements in the following three areas:

- *Absence of Constraints:* There is an assumption that a graph can continue to evolve indefinitely, unconstrained by any system wide or external resources. For most real world networks this is not the case. For example in communication networks every node in the network has a natural maximum connectivity. In the scale free model there is no such upper limit to node degree.

- *Fit to Real World Data:* The standard scale free model produces a degree distribution that follows a power law with exponent $\gamma = 3$. It is well understood that this is not an exact fit to real world data, which we highlight in Section 3. Many extensions exist that produce a better fit, some of which we survey later. It is clear that the degree distributions of real networks have more complex behavior than a simple fixed exponent power law.
- *Absence of a Physical Model:* The notion of scale freedom derives directly from the hypothesis of preferential attachment, that is in a dynamically evolving graph new nodes will more likely attach to nodes of higher degree. Whilst the scale free model provides a theoretical framework that points to high node degree making a node more likely to attract new connections, there is no fundamental explanation of *why* that should be so, and what physical processes may be at work that could produce that effect. It would be desirable if this could be explained using a first principles argument involving well understood mechanisms. This would further strengthen the fundamental premise of the scale free model.

In this paper we will attempt to address these challenges. We do so by proposing a simple extension to the standard scale free model, which introduces a hard cut off in the degree of a node, motivated by considerations from communications network design. This model has some attractive features, amongst which is a more accurate prediction of the power law exponent. Although extensions to the preferential attachment approach (most notably [5], [6] and [7]), can result in values of the power law exponent less than 3, we believe our model achieves this through a simple and natural extension to the traditional preferential attachment paradigm. Furthermore, as a consequence of introducing the constraint, we identify that the attachment probability introduces *superlinear* polynomial terms in node degree. This additional structure to the attachment probability is responsible for a richer scaling regime in node degree evolution. This structure allows us to compare in Section 4 both the constraints and scale free model to a novel model of evolution that argues from a stochastic perspective based upon recent developments in the structural entropy of a graph. By developing the outline of an entropic model we illustrate how both the standard scale free and our constrained model could be viewed as approximations to a more fundamental, statistical thermodynamic model of network growth.

In this section we will begin with a brief overview of the continuum analysis used in [1] to derive the principle results of scale free models, and at a very high level subsequent attempts to build upon and extend the model. We will make use of the same continuum approximation in our analysis.

We show in section 2 how the introduction of a simple environmental constraint into the scale free model can significantly improve its predictive power, and compare our *constrained* model to a range of more contemporary network data in section 3. As part of the verification of our *constrained* model, we also present results of simulations of network growth using our modified attachment probability defined in Section 2. An attractive feature of our extended model is that it reproduces the scale free model when we allow our constraint to tend to infinity. We are able to significantly outperform the ability of the scale free model to predict the exponent γ of the power law distribution across a wide range of real world data (results are summarized in Table 2). In particular for ten of the twenty three data sets analyzed (marked in Table 2 in bold) we are able to predict γ to within 10%, whereas the scale free model overestimates the value of γ by an average of 35% and in only four cases does it predict within the range 10-20%. Our constrained model therefore performs better than the standard scale free model on the first two issues identified above, but not on the third.

In Section 4 we propose a novel statistical thermodynamical (i.e. entropic) model of network growth. This addresses the third objective. Recent work on the behavior of communications networks by Tee *et al* [8,9] introduced a measure of the structural entropy of a node, derived from its degree and clustering coefficient. We show how this can lead to a direct derivation of scale free and constraint models, potentially explaining why scale freedom arises and why our constrained model is a better fit for networks as they grow and encounter connectivity limitations. We present in the same section some early results from numerical simulations of the entropic model, which show many of the features of the real world data we analyzed in Section 3.

1.2 The Scale Free Model

The Scale Free Model of Barabási, Albert and Jeong [3], [1] is based on two simple and fundamental assumptions:

- *Growth:* Starting with m_0 nodes and e_0 edges, we add a new node at each unit time step. When this node is added to the network, it connects to $m \ll m_0$ other nodes. This process continues indefinitely, such that after t unit time steps, there are $m_0 + t$ nodes, and $e_0 + mt$ edges. Eventually the constants in these expressions can be dropped as they are insignificant compared to t.
- *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.

Using a mean field theory approach the analysis explains both the power law scaling of real world networks [10], and the simultaneous resilience and vulnerability of networks to random and targeted attacks, respectively [11]. The approach taken in [3] begins by proposing the probability of a *randomly chosen node* i, capturing a connection to a new node, as solely dependent upon its degree k_i as:

$$\Pi_i = \frac{k_i}{\sum_j k_j} = \frac{k_i}{2mt} \ , \tag{1}$$

In the strictest sense the approximation $\sum_j k_j = 2mt$ should include the original nodes m_0 and their degrees, however for large values of t this can be effectively ignored, without loss of generality, as $2e_0 \ll 2mt$. By taking the continuous approximation, this naturally leads to the following ordinary differential equation for the time evolution of node i's degree $k_i(t)$:

$$\frac{dk_i(t)}{dt} = m\Pi_i = \frac{k_i(t)}{2t} \ . \tag{2}$$

Equation (2), can be solved subject to an initial condition that at time t_i, when node i is added, its degree $k_i = m$ to yield:

$$k_i(t) = m\left(\frac{t}{t_i}\right)^{1/2} \ . \tag{3}$$

In order to derive the degree distribution begin by assuming that t is fixed. At this stage the probability that $k_i(t)$ is smaller than a given degree k is:

$$P(k_i(t) < k) = P\left(t_i > \frac{m^2 t}{k^2}\right) = 1 - P\left(t_i \leq \frac{m^2 t}{k^2}\right) \ .$$

Developing the mean field approach we note that the ith node was chosen at random, so its time of introduction into the network t_i is a random variable. Given that nodes are added at each time step, the range of possible values for t_i are $1, 2, \ldots, (m_0 + t)$, and each value can occur with probability $\frac{1}{(m_0+t)}$. We can conclude that the random variable t_i is uniformly distributed and can write the probability of choosing a node i with a t_i smaller than $\frac{m^2 t}{k^2}$ as:

$$P\left(t_i \leq \frac{m^2 t}{k^2}\right) = \frac{1}{(m_0 + t)} \times \frac{m^2 t}{k^2} \ .$$

We can now state that the probability of a node having degree $k < k_i$ as:

$$P(k_i(t) < k) = 1 - \frac{m^2 t}{k^2(m_0 + t)} = \int_m^k P(x)dx \ ,$$

implying that $P(k) = \frac{\partial P(k_i(t)<k)}{\partial k}$, yielding the principal result of the Barabási-Albert Scale Free model:

$$P(k) = \frac{2m^2 t}{m_0 + t}\frac{1}{k^3} \ . \tag{4}$$

This predicts that on a log/log scale the slope of the degree distribution γ is identically 3. The result has been compared against many real world networks, and indeed the power law behavior has been seen in many examples and is one of the triumphs of the scale free model. The model, however, generally overestimates the value of γ and cannot explain the non linear behavior of the degree distribution at high values of k (as outlined in [12]). Reproduced in Table 1 from the data in [1] are some key parameters from a selection of the analyzed real world networks. The data is taken from a wide range of sources, which we supplement in Section 3, including the classic movie actor collaboration network from IMDB, a physical communications network, a biological network and a number of collaboration networks. A striking feature of all of these networks is both a limit to the degree of a node, and also that the value of γ is significantly lower than predicted by the scale free model (γ is calculated as described in Section 3.1.). Recent work [13] has highlighted a number of deficiencies in the scale free model, including deviations from the scale free degree distributions and the presence of cut offs in the maximum degree. It must be stated however that the model is strikingly powerful in its ability from a simple set of assumptions to explain many features of complex networks, from their small world property to the absence of a 'scale' in the degree distributions. This simplicity is powerful and hints at fundamental processes underlying the dynamics of network evolution.

Failure to capture the detail of the degree distributions of real world networks however, indicates that this simplicity must be supplemented with additional facets to the model of node attachment. In addition the appeal to node degree being the primary determinant of attachment probability is a modeling assumption and does not explain *why* that is the case. The principal argument is based on the concept of "the rich get richer", which is an equivalent statement to equation (1). In our view this is not a 'first principles argument', based upon fundamental physics. Given the success of the model and widespread acceptance of its validity and application in many fields from genetics to network design, it would be satisfying to link the derivation of equation (1) to core principles of physics. In this paper we start by exploring a next degree of approximation to the model to identify how environmental influences such as the presence of a top constraint for node degree alter the form of equation (1). In the model we propose this yields polynomial terms in k, which we hypothesize may be part of a series of corrections to the attachment probability. Using arguments based upon applying ensemble statistical mechanics to the entropy of a network vertex, we then propose an entropic model which naturally produces the concept of preferential attachment and constraints, and hints at further structure to the form of attachment probability in equation (1).

Table 1: Degree Distribution Parameters of some Real Networks [1]

Source	$\langle k \rangle$	**Max Degree**	γ
IMDB Movie Actors	28.78/127.33[1]	900	2.3
Internet Router	2.57	30	2.48
Metabolic, *E. coli*	7.4	110	2.2
Co-authors, SPIRES	173	1100	1.2
Co-authors, neuro	11.54	400	2.1
Co-authors, math	3.9	120	2.5

1.3 Extensions to the Scale Free Model

Before embarking on an investigation of our model, it is important to stress that many proposals to extend preferential attachment have been advanced. These alternative models to preferential attachment rely upon modifications to the probability of attachment beyond simple dependence on the degree of the node. The extensions range from ecologically inspired models such as the competition based approach of D'Souza in [15], to direct alterations of the form of equation (1) by introducing 'super-linear' terms in k, that is arbitrary powers of k. The model of Krapivsky *et al* [7], explicitly explores forms of attachment probability where the term in k is replaced by an exponential form k^{α}, where the exponent α can vary in the range $0 < \alpha < \infty$. By varying α it is possible to and produce very different forms of the degree distribution. These range from stretched exponential degree distribution to a super-linear zone for $\alpha > 2$ where one node captures a connection to all other nodes. In other work, notably Dorogovtsev *et al* [5], the concept of initial attractiveness of a node is introduced, which permits values of the power law exponent to vary and produces values of γ that are between $2 < \gamma < 3$. These models depend upon the concept of some nodes starting with a higher initial attractiveness than others in their ability to gain connections to new nodes. In some ways this is the opposite approach to the constrained model we propose in this paper, where nodes become progressively less attractive as they acquire connections and approach their limit.

It is perhaps the ecological, and physically inspired extensions that are most attractive alternatives to preferential attachment. We have already mentioned the competition based model of D'Souza [15] that uses an optimization approach in which the minimization of a cost function upon every node addition is used to determine which node the new node attaches to. This model produces an exponentially corrected degree distribution of the form $P(k) \propto k^{-\gamma} e^{-\alpha k}$. This degree distribution is similar to that which we see in the data analyzed in Section 3, and is an encouraging advance on the original preferential attachment model.

Another widely accepted approach, which builds upon the work of Dorogovtsev, was developed by Barabási in collaboration with Bianconi, This model parametrizes the attractiveness of the node using a *fitness* measure, η_i, and was introduced in [6], [16] and further developed in the work of Moriano *et al* [17], and Su *et al* [18].

The extended model proposes that the probability of attachment is modified to include the fitness parameter in the most general sense, as follows:

$$\Pi_i = \frac{\eta_i k_i}{\sum_j \eta_j k_j} \ . \tag{5}$$

To prevent this model requiring as many independent variables as there are nodes, the attractiveness η is fixed, or quenched, at node addition and is randomly assigned from an assumed probability distribution $\rho(\eta)$ for the parameter. The model permits an analogy between the graph and the Bose-Einstein treatment of ideal gases. This analogy relies upon the identification of a node vertex with an energy level of the gas ϵ_i, with the degree corresponding to the occupancy number of the energy level. Derivation of graph properties from statistical mechanical arguments is long established, including in the work of Newman and Park on exponential random graphs described in [19]. In the Bianconi-Barabási model the fitness parameter is defined as $\epsilon_i = -\frac{1}{\beta} \log \eta_i$, with β being identified as classical inverse thermodynamic temperature. The denominator of equation (5) is then easily identified with the partition function Z, familiar from the Bose-Einstein model of statistical mechanics. Using the probability distribution $\rho(\eta)$ of the nodes' fitness parameter as outline in [6], $P(k)$ can be analytically solved for in the case of the uniform distribution to yield:

$$P(k) \sim \frac{k^{-1+C}}{\log(k)}, \text{ where } C \text{ is a constant} \tag{6}$$

This model is attractive, and indeed does provide a closer fit to the data, including the presence of a cut-off on the maximum degree of a node.

The models described thus far all share a similar set up to the original preferential attachment mechanism, in that they consider a stepwise addition of a single node which connects to a variable number of pre-existing nodes. In recent work by Bianconi *et al*, this has been generalized to investigate models based upon the addition of simplicial complexes to a network rather than nodes as described in [20,21]. These models, referred to as Network Geometry with Flavor (NGF), introduce the concept of a d dimensional simplex, which is a fully connected clique of $d+1$ nodes. When $d = 1$ the model reduces down to the Bianconi-Barabási model, but higher dimensional simplices are hypothesized to more correctly represent the growth of networks where the unit of addition is a clique, such as a citation network being built from sub networks of frequently collaborating authors. The NGF model proceeds by adding a single node and links, so as to produce a new d dimensional simplex in the graph, by attaching the simplex to a randomly chosen $d-1$ existing face in the graph, governed by a generalized form of equation (5). The attachment probability is further parameterized by a flavor variable s which can take the values of $-1, 0, 1$ that allows the introduction of a generalized degree which counts the number of d dimensional simplices incident to a node. The range of flavor ensures that the form of attachment probability, which is beyond the scope of this survey to outline, produces a well behaved probability. The survey in [20] has a full and complete overview of the model. The attraction of these models is the generation of a rich set of possible graph geometries, including scale free, Apollonian and a form of graph deeply analogous to the form of graphs proposed in a range of approaches to Quantum Gravity.

Together with the competition model of D'Souza these more physically and ecologically inspired models provide motivation to explore other analogies with such processes to improve upon the standard preferential attachment. It

would be a significant insight if we could explain the experimental data based upon solely intrinsic properties of the graph such as node degree and local clustering coefficient of a node, with reference to how these relate to fundamental properties such as entropy and constraints. In the next section we propose an extension, based upon the concept of constraints to the maximum degree of a node. This constraint is motivated from real world concerns in many networks. For example in communications networks the number of physical connections a node can maintain has a hard limit, and even in social networks building a network of friends is subject to constraints of time and physical space. In Section 4 we show how both constraints and non-linear preferential attachment could arise from a deeper, more fundamental, entropic model.

2 A Pure Constraint Based Model

A core assumption of the scale free model is that new nodes attach to other nodes with a probability that is determined only by the degree of the target node; no other factors affect Π_i and attachment is unconditional. In most networks though this is not a fully accurate assumption, as most nodes will have some inherent upper limit on their capability to establish connections. We can imagine a network comprised of nodes capable of maintaining a maximum of c connections, with $c_i(t)$ being the point in time capacity of node i at time t. To simplify the treatment we assume the capacity of all nodes is equal across the network. In this case we could imagine modifying the probability of attachment to account for the nodes capacity as they accumulate connections, with a multiplicative factor to the preferential attachment probability Π_i. This assumption of uniform maximum capacity is an approximation that we justify by the simplicity of the theoretical analysis it permits. We seek to avoid introducing a family of free parameters, which would equate to a family of constraints, to preserve the theoretical elegance of the treatment. When we come to compare our *constrained* model to real world data it does require us to make reasonable estimates for the effective average constraint. We assume that this acts as a scaling factor for the attachment probability, similarly to the fitness factor introduced in the Barabási-Bianconi model [6],[16], in essence acting like a conditioning of the probability of attachment with the probability the node can accept the connection. In the most general sense, we can write this as the ratio of the nodes capacity relative to the time varying, average capacity of an arbitrary node, $\langle c(t)\rangle$ as:

$$\Pi_i^c = \zeta_i \times \Pi_i, \text{ where } \zeta_i = \frac{(c - k_i(t))}{\langle c(t)\rangle} \text{ and } \Pi_i = \frac{k_i(t)}{2mt} \tag{7}$$

To calculate $\langle c(t)\rangle$, we observe that at any time t a given node i will have an expected value of capacity $\langle c_i(t)\rangle = \langle c - k_i(t)\rangle$. As we assume that c is a shared maximum capacity across all nodes this reduces to $\langle c_i(t)\rangle = c - \langle k_i(t)\rangle$, and we note that $\langle k_i(t)\rangle$ is the expected value of a node's degree $k_i = \langle k_i\rangle$, which will be useful in section 3 when we will compare our constrained model against real networks. We can also estimate the expected value of the capacity of a node, by assuming a base uniform distribution of attachments in the absence of preference. After n nodes have been added, we will have added nc capacity to the graph, and consumed $2nm$ connections. In the simplest case for the average capacity of a node, after adding a large number of nodes n, we note that the average capacity must evolve to a constant as following:

$$\langle c(t)\rangle = \frac{nc - n2m}{n} = c - 2m \ . \tag{8}$$

Unfortunately as written this attachment probability is not sufficient as $\sum_i \Pi_i^c \neq 1$. This can be demonstrated by expanding Equation (7) as follows:

$$\sum_i \Pi_i^c = \frac{1}{(c-2m)2mt} \sum_i (c - k_i(t))k_i(t),$$
$$= \frac{1}{(c-2m)2mt} \left\{ c2mt - \sum_i k_i(t)^2 \right\} .$$

If we define δ as

$$\delta = \frac{\sigma}{\sum_i k_i(t)} = \frac{\sigma}{2mt} \text{ where,} \tag{9}$$

$$\sigma = \sum_i k_i(t)^2 - \sum_i \langle k_i(t)\rangle^2 \tag{10}$$

the normalization sum becomes,

$$\sum_i \Pi_i^c = 1 - \frac{\delta}{c - 2m} \ .$$

In general δ could be a function of time and degree, but as an approximation in our model we treat it as a constant of the system. We test that assumption in the simulations presented later in this section, which indicate that it is valid to assume that δ eventually stabilizes to a constant as the network evolves. We run these simulations of network growth to mimic the parameters for a selection of the real network data we analyze. Investigation of models where δ is a function of time (and potentially k_i) is an current avenue of research, and the subject of future work. For our attachment probability to be a valid probability measure we need to establish that $\frac{\delta}{(c-2m)} \geq 0$ and that $\frac{\delta}{(c-2m)} \leq 1$. In the first instance the numerator of Equation (9), as defined in Equation (10), is the variance of k_i across the graph, and so is strictly positive. Providing that $c > 2m$, we can safely assume $\delta \geq 0$.

Regarding the upper limit of δ, we can appeal to Popviciu's inequality (see [22]) for a bounded distribution, with $k_{max} = c$ and $k_{min} = m$. This states:

$$\sigma \leq \frac{1}{4}(k_{max} - k_{min})^2 \leq \frac{1}{4}(c-m)^2 \ ,$$
$$\Rightarrow \frac{\delta}{(c-2m)} \leq \frac{(c-m)^2}{8(c-2m)mt} \ .$$

For times $t > \frac{(c-m)^2}{8m(c-2m)}$, we then conclude that as required $\frac{\delta}{(c-2m)} \leq 1$. With these limits established, we can modify the attachment probability by adding in δ to produce a form for the attachment probability, which sums to unity at each time step across all nodes, below:

$$\Pi_i^c = \zeta_i \times \Pi_i, \text{ where } \zeta_i = \frac{(c+\delta-k_i(t))}{c-2m} \text{ and } \Pi_i = \frac{k_i(t)}{2mt} \tag{11}$$

For convenience, we can further simplify the expression for ζ_i, as follows:

$$\zeta_i = \alpha\left(1 - \frac{k_i(t)}{(c+\delta)}\right),$$
$$\text{where } \alpha = \frac{c+\delta}{c-2m}, \text{ or equivalently } \alpha = \frac{c+\delta}{c-2\langle k_i\rangle} \ . \tag{12}$$

We can now write the complete probability of attachment as:

$$\Pi_i^c = \frac{\alpha k_i(t)(c+\delta-k_i(t))}{2m(c+\delta)t} \ . \tag{13}$$

For comparison with the Barabási-Albert model, using $\alpha = \frac{c+\delta}{(c-2m)}$ from equation (8) we can rewrite Π_i^c as follows:

$$\Pi_i^c = k_i(t)\frac{(\frac{c+\delta-k_i(t)}{c-2m})}{2mt} \approx k_i(t)\frac{1}{2mt}, \text{ for large } c.$$

This recovers the standard Barabási-Albert model in the case that the constraint c is infinite and therefore does not interfere with the dynamics of the network's evolution. Following the continuum approach, and dropping the explicit time dependency of k_i for clarity, we can substitute this into equation (2), to obtain

$$\frac{\partial k_i}{\partial t} = m\Pi_i^c = \frac{\alpha k_i(c+\delta-k_i)}{2(c+\delta)t} = \frac{\alpha k_i}{2t} - \frac{\alpha k_i^2}{2(c+\delta)t} \ , \tag{14}$$

with the fraction multiplied out for convenience later. This is directly solvable by separating as follows:

$$\frac{1}{\alpha}\int \frac{dk_i}{k_i(c+\delta-k_i)} = \frac{1}{\alpha(c+\delta)}\int \Big\{\frac{1}{k_i} + \frac{1}{c-k_i}\Big\}dk_i = \frac{1}{2(c+\delta)}\int \frac{dt}{t} \ ,$$

whose solution is:

$$\log\left(\frac{k_i}{c+\delta-k_i}\right) = \frac{\alpha}{2}\log(t) + \theta \ ,$$

or in simplified form

$$k_i = (c+\delta)e^{\theta}\left(\frac{t^{\alpha/2}}{e^{\theta}t^{\alpha/2}+1}\right) \ .$$

Following the continuum method in [1] we apply the initial condition that $k_i(t) = m$ at time $t = t_i$, to obtain:

$$k_i(t) = \left(\frac{\rho(c+\delta)\big(\frac{t}{t_i}\big)^{\alpha/2})}{1+\rho\big(\frac{t}{t_i}\big)^{\alpha/2}}\right), \quad \text{with } \rho \text{ defined as, } \rho = \frac{m}{c+\delta-m} \ . \tag{15}$$

Again, we note that as $c \to \infty$, $\rho(c+\delta) \to m$, $\alpha \to 1$, and so equation (6) reduces to

$$k_i(t) = m\left(\frac{t}{t_i}\right)^{1/2} \ ,$$

the standard result from the continuum analysis of Barabási and Albert [1],[3]. We then note that the probability that a node has degree $k_i(t) < k$ is:

$$P(k_i(t) < k) = P\Big(t_i > \frac{\rho^{2/\alpha}(c+\delta-k)^{2/\alpha}t}{k^{2/\alpha}}\Big) = 1 - P\Big(t_i \leq \frac{\rho^{2/\alpha}(c+\delta-k)^{2/\alpha}t}{k^{2/\alpha}}\Big) \ .$$

Assuming uniform probability for the choice of node introduction time t_i of $\frac{1}{(m_0+t)}$ we arrive at the expression:

$$P(k_i(t) < t) = 1 - \frac{\rho^{2/\alpha}(c+\delta-k)^{2/\alpha}t}{k^{2/\alpha}(m_0+t)} \ .$$

Although somewhat more complex than the expression in [1] it is nevertheless simple to compute the distribution equation $P(k) = \frac{\partial(k_i(t)<k)}{\partial k}$ to obtain the main result of our constrained model:

$$P(k) = \frac{2(c+\delta)\rho^{2/\alpha}t}{\alpha(t+m_0)}\left(\frac{(c+\delta-k)^{\frac{2}{\alpha}-1}}{k^{\frac{2}{\alpha}+1}}\right) \ . \tag{16}$$

In appendix A we examine the asymptotic behavior of Equation (16), which verifies that by careful manipulation the standard result of the scale free model $\gamma = 3$, is recovered in the limit $c \to \infty$. Further, this analysis also indicates that the dominant contribution to degree distribution for $k \ll (c+\delta)$, produces a scale free log linearity with power law exponent $\gamma = \frac{2}{\alpha} + 1$. This equivalence to a more straight forward power law, but with an exponent $\gamma < 3$ for values of $k \ll (c+\delta)$ indicates that the presence

of a constraint influences the behavior of our model even for nodes early in their evolution. This is a significant result and we make use of it to compare the predictions of our theory against real network data and simulations in section 3.

The result in equation (16) has some interesting implications, as the presence of a finite capacity c alters the scale factor for the distribution of the nodes, *whilst* preserving the essential aspects of scale free behavior. By way of example, the data for the IMDB movie actor database, as presented in Table 1, is plotted in Figure 1b, along with results from a simulation of our model. The movie actor database naturally produces a graph by assigning a vertex for each actor and connecting two vertices when the actors have acted in the same film. Figure 1b contains a theoretical plot of the distribution taken directly from equation (16), using $\langle k \rangle = 127$, $c = 900$ and with initial conditions of $m_0 = 100$, which we take from Table 1 . For this plot we set $\delta = 205$, which we take directly from the simulation, which we discuss in the next paragraph. The unmodified scale free model would give a value of γ of exactly 3, but our modification has an initial value of $\gamma = \frac{2}{\alpha} + 1$, which increases as $k \to c$ and reaches a limit when $k = c$. To calculate γ we can take $c = 900$ from the dataset in Table 1 and $k = 127.33$, with the estimated value of $\delta = 243$ (we average the ratio of δ to c), to yield $\gamma = 2.35$, versus the measured value of 2.3 in [1] and 2.43 from our simulation. By comparison, to the scale free model, our approach predicts the value of γ to 2.29%, compared to 30.4% for scale free, a significant improvement. In addition, there is no explanation in the scale free model for the degree of a node in the graph having a maximum value.

To further verify our model, and in particular the assumption that δ can be effectively treated as a constant, simulations were run using the form of preferential attachment probability in equation (13), for a network sharing the same parameters of maximum degree and average degree as the IMDB network. We present those results in Figure 1a. The simulation was run for a selection of initial parameters to asses the evolution of δ, and in each case the value quickly converges to a constant. Turning to the simulation of degree distribution, in Figure 1b the essential scale free nature of the network obtained is visible on the log scale graph, as is the goodness of fit and agreement between the simulation with a theoretical plot of $P(k)$ using the same simulation parameters. Using the techniques described in [23], we can measure γ, and obtain a value of 2.40 versus a calculated value from equation (16) of 2.41, which is in close agreement.

We also ran simulations for the Patents Citation graph (Figure 1c) and the Web Provider network (Figure 1d), which both produce similarly good results to the IMDB network in terms of the closeness of fit between the simulated and theoretically obtained $P(k)$. We can conclude that the constrained model is a good representation of networks with a simple maximum degree constraint.

Motivated by this example and simulation, in the following section we extend our analysis to a range of more recent, publicly available, network data to investigate further the accuracy of our constrained model.

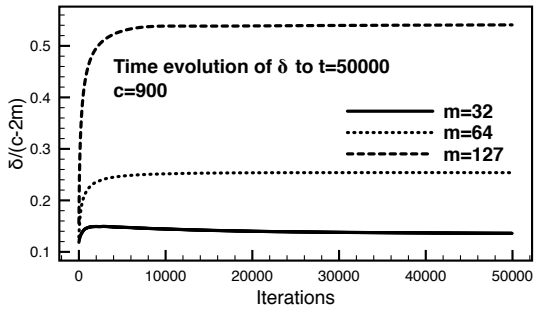
3 Analysis and Comparison of Constrained versus Preferential Attachment

3.1 Data and Methods

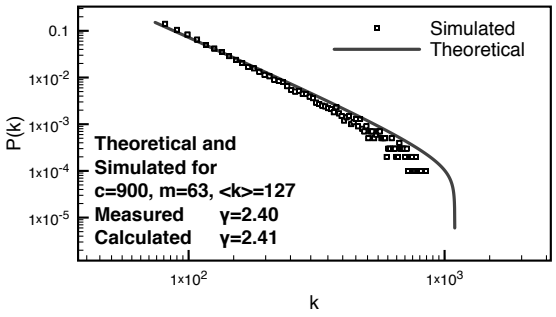
In this section we present the analysis of an extensive collection of network datasets comprising virtual, transport, and communications networks. The bulk of this data is publicly available through the Stanford Large Datasets Collection [24] which comprises an excellent repository of large graphs. The Twitter follower data is provided by [25], and the rest of the datasets are reproduced from publications such as [1], the Internet Topology Zoo [26]. We have one proprietary graph built from the topology taken from a large commercial deployment of network infrastructure used to deliver a top 10 Internet portal service (see [8]). The produced graphs fall into the following categories:

- *Social Networks.* These include Twitter, Facebook, Pokec graphs of the relationships between users. Typically each user is a node and nodes have links if the users have some form of relationship with each other. For example in the case of Twitter this relationship derives from one user 'following' another.
- *Collaboration and Citation Networks.* These cover a wide range of publicly available data, including the Arxiv citation, Patent Citation and co-authorship graphs as examples. Graphs are constructed by creating a vertex for each unique user or paper and then connecting the vertices if they share authorship with another vertex or directly cite it.
- *Communications Networks.* These networks, such as the Internet Router, IT Zoo, Web Provider and Berkeley Stanford Web Graph are constructed by representing physical or virtual nodes by a vertex in the graph and communications links as edges connecting the vertices.
- *Biological Networks.* These networks use a graph to represent a biological process, for example the metabolism of the *E. coli* organism. Nodes in the graph represent a molecule or intermediate state in the process used by *E. coli* to release energy from its food sources, with edges connecting nodes where a reaction or transition occurs. Similar networks exist for other biological processes (e.g. for the genetic cause and effect in cancers and disease epidemic spreads).

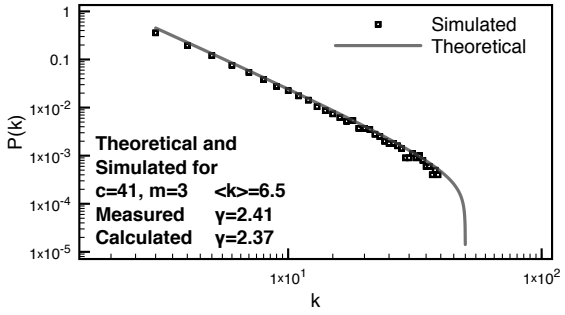
Analysis of the data was undertaken using a program and graph datastore which is available from the authors on request. The source data was often very large (the Twitter data contains for example over 10 million edges), and extracting values for the max degree and $\langle k \rangle$ is not necessarily evident. Some of the data had some extreme outliers in terms of node degree, and to avoid skewing the results, we estimated the constraint at the 99^{th} percentile of k rather than the maximum value in the data. This



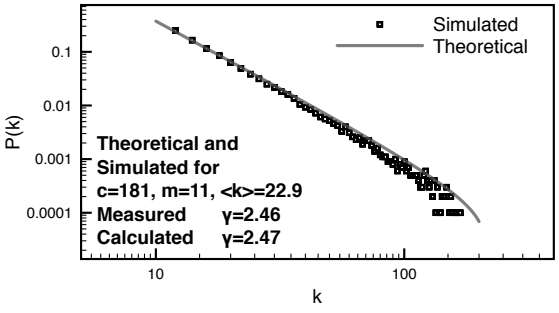
(a) Convergence of $\frac{\delta}{(c-2m)}$ over 50,000 Iterations in a Simulation of Constrained Attachment



(b) Simulation and Theoretical Degree Distribution using Equation(13) and IMDB Parameters at $t = 50,000$



(c) Simulation and Theoretical Degree Distribution using Equation(13) and Patents Parameters at $t = 50,000$



(d) Simulation and Theoretical Degree Distribution using Equation(13) and Web Provider Parameters at $t = 50,000$

Fig. 1: Simulation Results for Constrained Attachment

is consistent with the methodology taken in the theoretical analysis, where we made an assumption of the node degree constraint being constant for all nodes. This is a simplification, but one with great benefit in the analytical treatment of the model. The elimination of outliers at first sight may seem inconsistent with the assumption of a single constraint in the capacity of a node, but it is expected that the real world data will contain perhaps many different constraints, and that the average behavior of the graph will be most influenced by the effective maximum established at the 99^{th} percentile. Further, the data above the 99^{th} percentile in k is typically very sparse and may contain spurious data points, which this cut off eliminates. In Figure 2 we present the variation of the calculated value of γ with the choice of percentile at which to choose c. The range of calculated values as we move from the 98.2^{th} to the 100^{th} percentile is 2.20 to 2.69, a range of $\pm 9\%$ either side of the chosen value of $c = 41$. We believe this further strengthens our choice of the 99^{th} percentile as the appropriate cut off for measuring c.

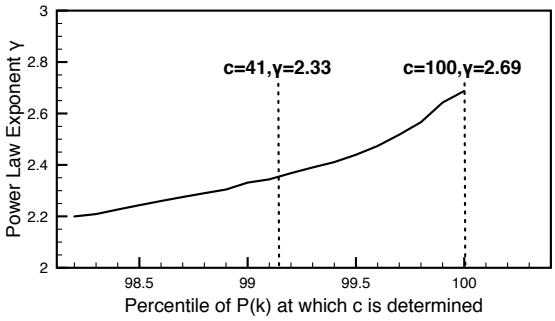


Fig. 2: Variation of Calculated Values of γ with Choice of Percentile for c for the Patents Graph

For $\langle k \rangle$ we require the expected value of the degree. This was calculated by computing the weighted mean, a discrete approximation of $\langle k \rangle$, which is truly only valid if k is a continuous variable. This is consistent with the approximation of continuity inherent in the continuum analysis approach.

To compare against the actual value of γ, power law exponent, we followed the techniques outlined in [23] to both asses the presence of a scale free distribution and obtain the value of γ. For the datasets we analyzed, which can be seen visually in Figures 3, 4 and 5, there is a considerable portion of the distribution which has a well defined straight line on the log/log plots, illustrating the intrinsic power law distribution of node degree. We capture the measured values of these power law exponents in Table 2.

3.2 Analysis

In the summary Table 2 it is compelling to note that in all but a few cases the constrained model is more accu-

Table 2: Comparison of γ Predictions Between Preferential Attachment and Constraints Model

Source	$< k >$	c	γ Calculated	γ Measured	Δ **Constraints**	Δ Scale Free
Patent Citation[2]	6.57	41	2.33	2.31	**0.75%**	29.66%
IT Zoo[1]	2.26	10	2.32	2.36	**1.63%**	27.19%
Internet Router[7]	2.57	30	2.44	2.48	**1.64%**	20.97%
Arxiv - Condensed Matter[2]	9.13	51	2.32	2.37	**2.18%**	26.39%
IMDB Movie Actors [7]	127.33	900	2.35	2.30	**2.29%**	30.43%
Pokec[3]	39.27	180	2.30	2.25	**2.29%**	33.34%
Airport Connections[9]	11.18	126	2.35	2.29	**2.82%**	31.19%
Arxiv - HepTh (Cit) [2]	26.75	165	2.34	2.44	**3.82%**	23.10%
Twitter (Circles)[4]	33.94	264	2.37	2.47	**3.90%**	21.43%
Arxiv - HepTh (Collab)[2]	22.05	285	2.37	2.51	**5.37%**	19.67%
Web Provider[5]	4.18	36	2.09	2.23	**6.33%**	34.48%
Co-authors, math[7]	3.90	400	2.69	2.5	**7.60%**	20.00%
Berkeley Stanford Web[6]	24.59	173	2.31	2.35	10.45%	27.74%
Metabolic, *E. coli*[7]	53.51	137	2.47	2.20	12.20%	36.36%
AS Skitter[2]	54.13	150	1.94	2.34	17.27%	28.14%
Facebook[4]	42.99	198	2.25	2.75	18.08%	9.23%
Arxiv - Astro Phys[2]	23.81	144	2.33	2.87	18.70%	4.61%
Co-authors, neuro[7]	11.54	400	2.53	2.1	20.42%	42.86%
Enron Email[6]	40.25	280	1.84	2.42	23.87%	24.02%
Twitter (Follower)[8]	8.63	90	1.56	2.39	34.77%	25.46%
PA Road Network[6]	5.41	9	1.71	2.69	36.27%	11.71%
Co-authors, SPIRES [7]	173.00	1100	2.69	1.2	124.17%	150.00%

rate in its predictions of γ than the standard scale free model. Indeed in the case of the Patent Citation, Internet Topology Zoo, Pokec, the real world network from a Web Provider, and a number of the citation networks and social networks, it comes very close to an exact prediction. Given that the motivation to investigate the constrained model originated from considerations of network design in communications networks, it is interesting to see that this has some strong applicability to non-physical networks.

We also present the analysis both as a collection of log/log distribution graphs in Figures 3, 4 and 5 and also summarize the key prediction of γ against the standard value of 3.0 from preferential attachment in Table 2. In the log/log plots we overlay the value of c at 99^{th} percentile, the average value of γ to this constraint and the expected value of the node degree $\langle k \rangle$. In each of Figures 3, 4 and 5, we also overlay the theoretical prediction for the distribution $P(k)$ obtained by substituting the values of γ from Table 2 into Equation (16). The agreement between the predicted values of γ and the measured ones for our datasets is evident from these combined theoretical and experimental plots, at least for portions of the distribution. A consequence of the selection of c at the 99^{th} percentile is that our theoretical curve displays a cut off earlier than the experimental data, which is to be expected.

The striking feature of many of the degree distributions is the absence of strict linearity, contrary to the predictions of the standard scale free model, and also the marked increase in γ at high values of k, a key prediction of our constrained model and a necessary precursor to a hard constraint in the value of k. In the social network data we analyzed this is best illustrated in Figures 3a, 3c and 3b. Similar behavior is also present in the citation network (perhaps the best example being Figure 4d), and again in the infrastructure graphs, particularly the Internet Topology Zoo (Figure 5a). It is interesting to speculate what the nature of the constraint is in the social networks, but this is perhaps explained by the effective limitations, no matter how small, on the amount of time people can feasibly spend on social networking platforms. Indeed in almost every conceivable network a constraint is a natural feature. Whether the node in the graph is a physical device, and individual engaged in an activity such as writing papers, or web site hyperlinks, there is a limitation to the connections a node can have. In some cases these are hard design limits such as ports on a network switch, in others it is simply the capacity of a human being, with a fixed lifespan, to blog, interact, star in a movie or engage in any other social activity. In every case our experimental data bears this out.

In the following Section 4 we point out how the two models may well be related to a fundamental dynamical principle that arises from thermodynamic considerations of network evolution. Critically this analysis derives the form of pref-

erential attachment presented as an axiom in the scale free model.

4 Dynamical Evolution of Scale Freedom

In our treatment thus far we have followed the continuum model of Barabási-Albert with the addition of a constraint-based factor to the attachment probability. However, we can attack the problem from a more fundamental viewpoint. Essentially, we argue that the evolution of a graph satisfies the criteria for a treatment based upon considerations of entropy from a statistical mechanics perspective, in accordance with the 2^{nd} law of thermodynamics. In any isolated physical system the entropy of the system will tend to a maximum unless energy is input to prevent that. For a classic treatment see [32]. In natural processes this tendency to increase entropy can be modeled as a macroscopic force on the system. This entropic force is responsible for both the elasticity of certain polymers and the biological process of osmosis. Indeed if thermodynamic temperature is written as $\mathbf{T}$ and entropy $\mathbf{S}$, one can state the entropic force F acting on a body when a process changes entropy as follows:

$$F = \mathbf{T}\Delta\mathbf{S} \ . \tag{17}$$

To begin our treatment of graph evolution from fundamental thermodynamic principles, it suffices to pose the problem in an appropriate manner. Consider an existing graph of m_0 nodes and e_0 edges in thermal equilibrium with an infinite supply of unattached nodes, each capable of connecting to m nodes in the event that it comes into contact with the existing graph. At every time-step we imagine that such an interaction occurs and the new node connects to m others. Our problem is to identify the probability of attachment for a node according to its degree k, and thus derive the degree distribution. More strictly, it is necessary to consider an ensemble of all possible graph configurations, *at every time step*, to enable statistical treatment of this process. This requirement to consider an ensemble of configurations is at first sight an added complication, but in fact is critical in permitting the analysis of the model. Whenever we consider a randomly selected node, for example in equation (18), it is important to recognize that we must average any interaction with the remaining graph over *all possible graphs* that can be constructed from the subgraph obtained by removing the randomly selected node and all edges connected to it. This ensemble average is further constrained by the total number of vertices and edges being unchanged after the removal of the random node. This requirement to average over all possible graph configurations at each time step justifies the approximation we make to calculate, for example, the average clustering coefficient.

The probability of attachment to a random node must statistically and universally seek to maximize total entropy. Our model proposes that the probability of this random node acquiring new links is a result of the relative strength of the entropic force of attachment to the randomly chosen node versus any other node in the graph. Those nodes which exert the highest entropic force relative to the rest of the nodes in the network will gain the most links, and we write this mathematically as:

$$\Pi_i = \frac{F(v_i)}{\sum_{j\neq i} F(v_j)} \tag{18}$$

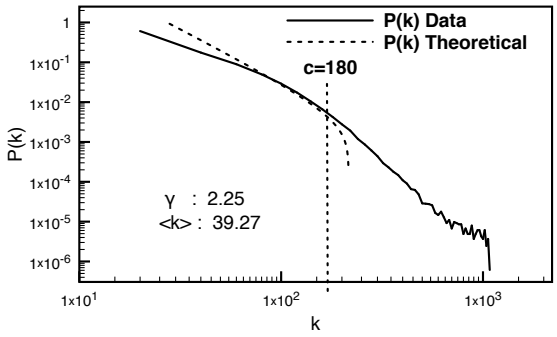
where $F(v_i)$ is the entropic force of attraction to node i. This expression governs the individual interaction that our randomly selected node has with a particular graph configuration, analogous to the elastic collision equations used to formulate the statistical treatment of ideal gases. In a similar way we cannot easily analytically formulate the dynamical equations of the graph from this equation as they are very large, and so to derive the degree evolution equations from this formulation we utilize statistical ensemble arguments. Considering all possible configurations of the graph $G(V(t), E(t))$ at a fixed time t, the denominator of equation (18) is computed as an expectation value of the relative force of attaching to any other node, across all possible graphs at time t in the ensemble that our random node could be connected to. At a given time t in the evolution of the graph the numbers of vertices $|V(t)|$ and edges $|E(t)|$ are constant, but we do have to consider all possible graph configurations of that number of vertices and edges. This will ultimately change the average of the change in entropy that the node could make on connecting to any other node in the graph other than our randomly selected node v_i. In this way we collapse the denominator to the expected value of this entropy change, averaged across all possible connection points in all possible members of the ensemble. We write this as $\mathbf{T}\times|V|\times\mathbb{E}(\Delta\mathbf{S})$. As the graph becomes larger, we make the assumption that the value of $|V|\times\mathbb{E}(\Delta\mathbf{S})$ is effectively constant, and factor this out. We base this assumption on the fact that most real world networks do indeed demonstrate some form of steep drop in the distribution of node degrees, so that the vast majority of nodes posses low degree (an important claim of [4] and [1]). It seems reasonable to assume that with such a restricted degree sequence most nodes will contribute a similar amount to the change in entropy, and this expected value will stabilize to a constant. More complex analysis could admit a time varying value of this constant, as strictly both V and $\mathbb{E}(\Delta\mathbf{S})$ may have complex time dependence, but for simplicity we assume:

$$\epsilon = \frac{1}{|V|\times\mathbb{E}(\Delta\mathbf{S})} \ .$$

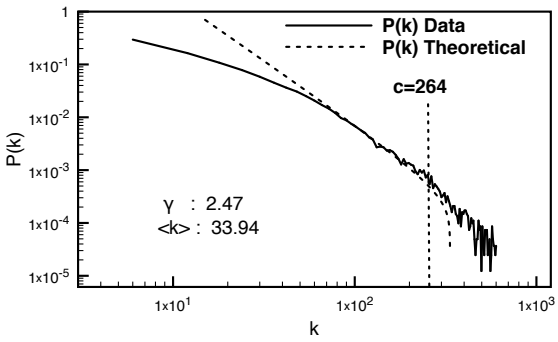
With this assumption equation (18) simplifies and $\mathbf{T}$ factors out to yield

$$\Pi_i = \epsilon\Delta\mathbf{S}_i \ . \tag{19}$$

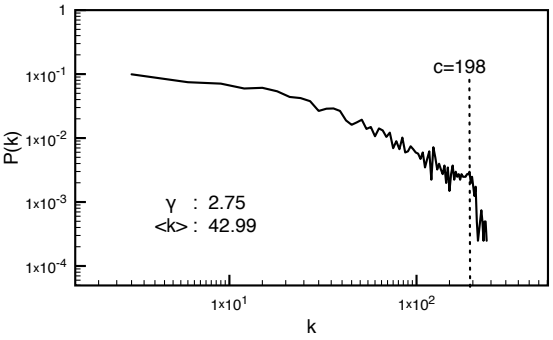
In general $\mathbf{S}_i$ is a function of potentially many variables x_i, but certainly depends upon k_i and time t. We can calculate $\Delta\mathbf{S}_i$ as a total differential, $\Delta\mathbf{S}_i(x_j) = \sum_{x_j} \frac{\partial\mathbf{S}_i}{\partial x_j}\Delta x_j$, but we can assume for simplicity that t is fixed and the dependence is purely upon k_i. In this case $\Delta\mathbf{S}_i = \frac{d\mathbf{S}_i}{dk_i}\times\Delta k_i$,



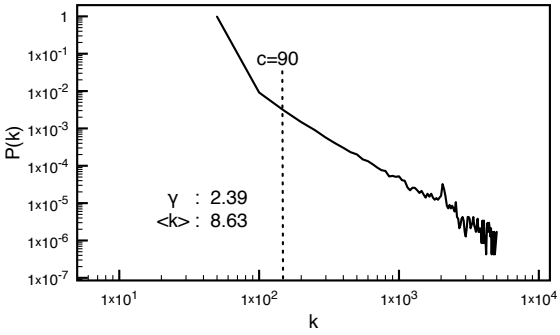
(a) Pokec - Slovakian Social network Friendship Graph, Theoretical and Experimental [28]



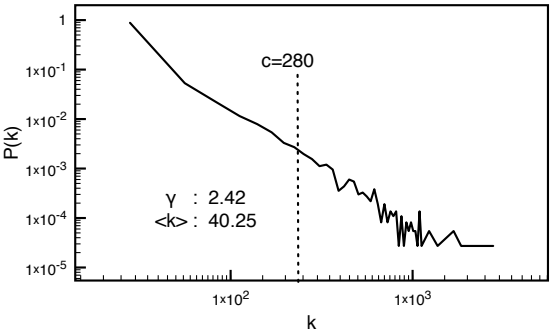
(b) Twitter Friendship Circles, Theoretical and Experimental [29]



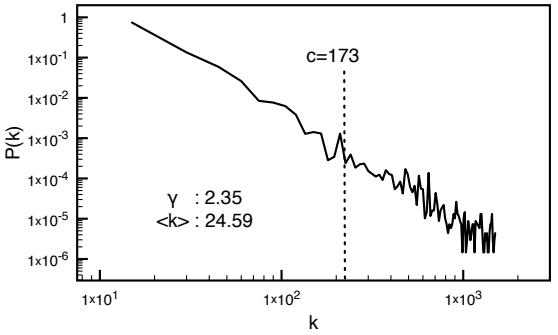
(c) Facebook Friendship Network [29]



(d) Twitter Follower Network [25]



(e) Enron Email Communication Network [30]



(f) Berkley Stanford Web Interconnection Network [30]

Fig. 3: Degree Distributions from Social Networking and Web Networks on a Logarithmic Scale

with, for a single time step, $\Delta k_i = 2m$. This gives us our expression for attachment probability:

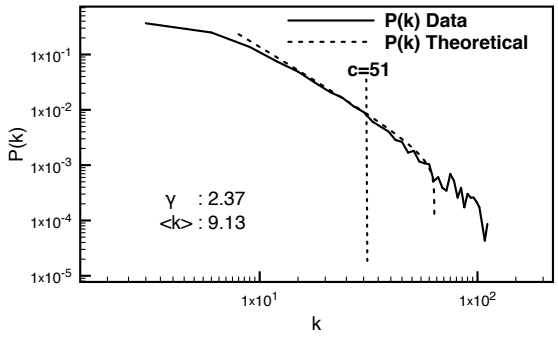
$$\Pi_i = \epsilon 2m \frac{d\mathbf{S}_i(k_i)}{dk_i} \ . \tag{20}$$

To make use of equation (20) we require an expression for the entropy of a node in the graph. The subject of the entropy of a graph has a long history, originating in the work of Körner on the informational entropy of signals described in [33] and [34]. Many approaches to calculating the entropy of a graph have been proposed, including the use of the eigenvalues of the adjacency matrix (see [35], and ensembles of networks with similar degree sequences (proposed in [36]). Unfortunately these concepts relate to the global value of entropy for a graph, and do not have utility when calculating the change in entropy as a new node connects.

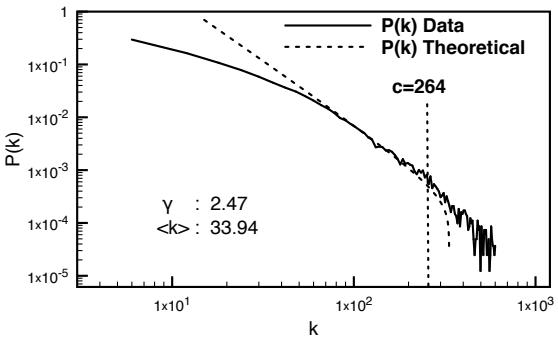
A series of papers by Dehmer ([37],[38]) formalized the concept of the individual entropy of a node. In recent work [8] we built upon this formulation to define a local vertex measure (referred to in [8] as NVE', and equivalent to our definition of $\mathbf{S}_i$ here) in terms of its relative degree as:

$$\mathbf{S}_i(k_i, t) = \frac{1}{C_i^1} \times \frac{k_i}{2|E(t)|} \log \frac{2|E(t)|}{k_i} \ , \tag{21}$$

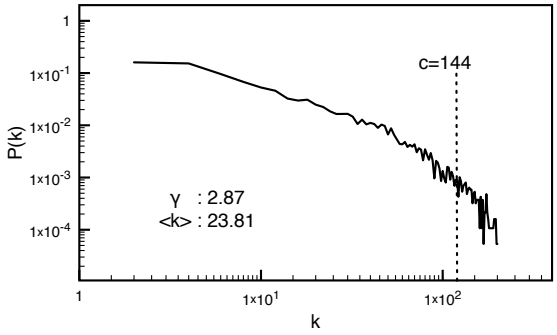
where C_i^1 represents a modified clustering coefficient of the 1-hop neighborhood of the node v_i. Contrary to the more common point-deleted neighborhood clustering coefficient, C_i^1 preserves the node in the calculation to measure similarity to the local perfect graph K_n of order $n = k_i+1$.



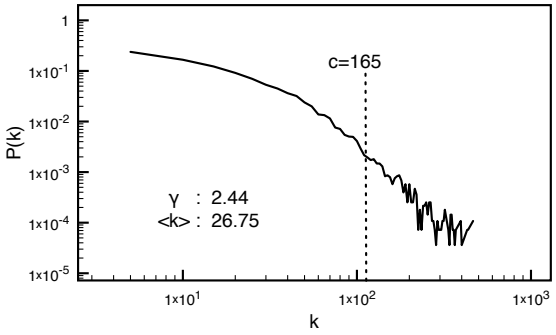
(a) Arxiv Condensed Matter Citation Network, Theoretical and Experimental [27]



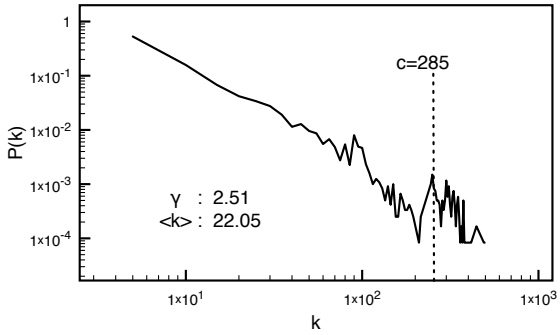
(b) Patent Citation Network, Theoretical and Experimental [27]



(c) Arxiv Astro-Physics Citation Network [27]



(d) Arxiv High Energy Physics Citation Network [27]



(e) Arxiv High Energy Physics Collaboration Network [27]

Fig. 4: Degree Distributions from Collaboration and Citation Networks on a Logarithmic Scale

For convenience we give an explicit definition of the 1-hop neighborhood N_i^1:

$$N_i^1 = \{v \in V \mid d(v_i, v) \leq 1\} \cup \{v_i\} \ ,$$

and the related '1-edges' E_i^1 as

$$E_i^1 = \{e_{jk} \in E \mid v_j \in N_i^1 \text{ and } v_k \in N_i^1\} \ .$$

We can then define the modified clustering coefficient to be

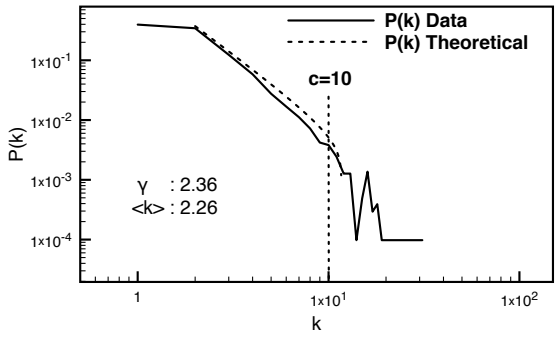
$$C_i^1 = \frac{2|E_i^1|}{k_i(k_i+1)} \ . \tag{22}$$

At this point we can make use of the fact that we must consider all possible intermediate graph configurations to assume effective uniformity in the graph to calculate $|E_i^1|$, and assert that for a given node, $|E_i^1| = \frac{k_i+1}{|V|} \times |E(t)|$. This then yields for the clustering coefficient the following expression:

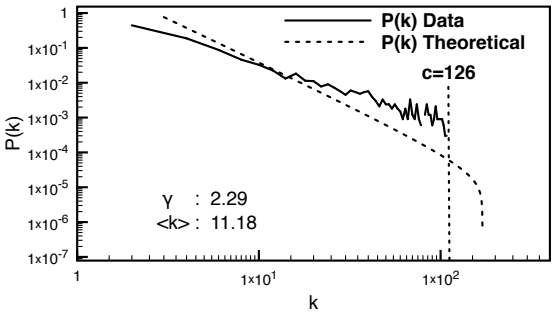
$$C_i^1 = \frac{2|E(t)|}{k_i|V(t)|} \ . \tag{23}$$

Given that at every time-step we add one node to the graph, connecting to m other nodes we can write $|V| = m_0 + t$, and $|E| = e_0 + mt$. In general as the model evolves, $t \gg m_0$ and similarly, $mt \gg e_0$, these simplify to $|V| = t$ and $|E| = mt$. Substituting back in we obtain the following equation for vertex entropy at v_i at time t as:

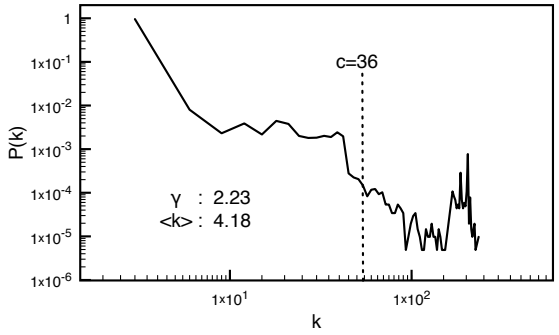
$$\mathbf{S}_i(k_i, t) = \frac{k_i^2}{4m^2 t} \log\left(\frac{2mt}{k_i}\right) \ . \tag{24}$$



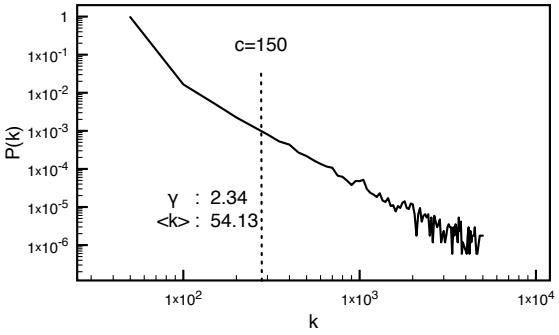
(a) Internet Topology Zoo Network, Theoretical and Experimental [26]



(b) Airport Flight Interconnection Network, Theoretical and Experimental [31]



(c) Web Provider Datacenter Network [8]



(d) Internet Autonomous Systems Network [27]

Fig. 5: Degree Distributions from Infrastructure and Communications Networks on a Logarithmic Scale

In the analysis undertaken by Tee *et al* in [8,9], this quantity was identified as sharing some of the properties of the structural entropy of the graph when summed across all vertices. In particular, the extremal behavior of the summed vertex entropy was proven to be minimized by the perfect graph of order n, K_n, and maximized by the star graph of order n, S_n, for simply connected undirected graphs. From the perspective of dynamical evolution of networks, this is consistent with the approach in our analysis. The perfect graph K_n will tend towards a more node level disordered graph such as S_n as addition of nodes selects targets such as to increase the value of $\mathbf{S}_i$ in Equation (24). From a purely statistical mechanics perspective one can consider each connected graph on n nodes and $|E|$ edges as representing a micro-state. The perfect graph is achievable in precisely one unique configuration if edges are indistinguishable, whereas other configurations, S_n for example, can be achieved by selecting any one of the nodes as the hub vertex. In this way the result that increases in entropy tends to destroy cliques and regular ordered graphs is consistent. From this perspective we would expect dynamic processes to favor the attachment to nodes where the increase in $\mathbf{S}_i$ is greatest. From here it is straightforward to follow through the continuum analysis as described in [1]. For the time evolution of k the following equation, is obtained:

$$\frac{dk_i}{dt} = 2m\Pi_i = -\epsilon \frac{k_i}{t}\left\{\frac{1}{2} + \log\left(\frac{k_i}{2mt}\right)\right\}. \qquad (25)$$

Although at first sight this nonlinear ODE appears intractable, in fact an analytic solution is available. Making the change of variables $y = \log k$ and $x = \log t$, so that $\frac{dy}{dx} = \frac{t}{k}\frac{dk}{dt}$, we see that (25) becomes

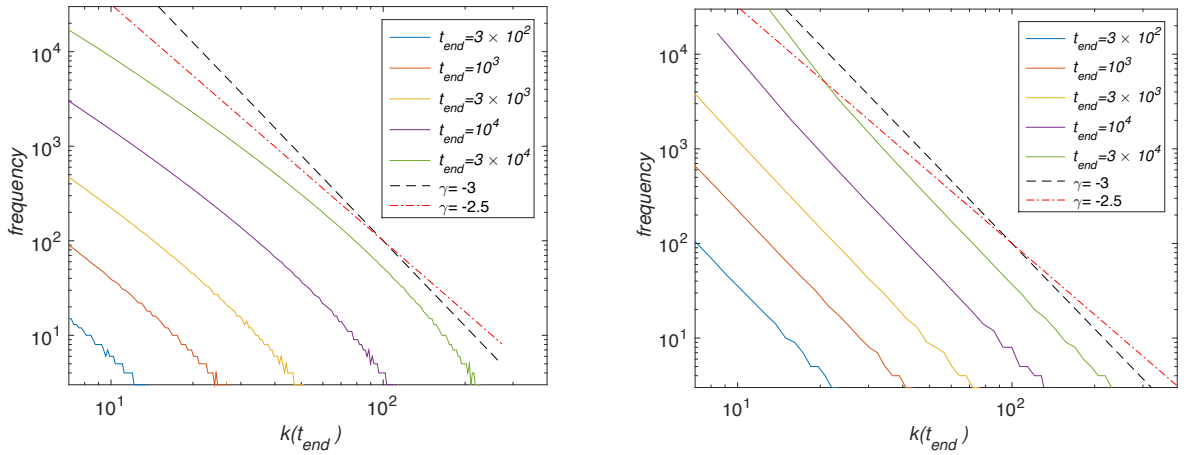
$$\frac{dy}{dx} = -\epsilon\left[\frac{1}{2} + y - \log(2m) - x\right]$$

This is now a linear ODE which can be solved by standard methods. Applying the initial condition $k_i(t_i) = m$ the solution is found to be most conveniently expressed in the form

$$\log k_i(t) = \log(2mt) - \frac{1}{2} - \frac{1}{\epsilon} + \left[\frac{1}{2} + \frac{1}{\epsilon} - \log(2t_i)\right]\left(\frac{t_i}{t}\right)^{\epsilon} \qquad (26)$$

For values of $\epsilon < 1$ the behavior of $k_i(t)$ is similar to the Barabasi–Albert model: degrees increase monotonically but at an ever decreasing rate. An analytic form for the degree distribution, analogous to (3) does not seem straightforward to derive.

Figure 6 compares numerically computed degree distributions from the model (26) (shown in figure 6a) and the Barabasi–Albert model, shown in figure 6b. In each case a new node was added to the network every 0.5 time units, setting $m = 5$ and growing the degrees of existing nodes according to (26) or (3) respectively. Degree distributions are plotted for fixed end times t_{end}, taking the



(a) (b)

Fig. 6: Degree distributions for growing networks at fixed end times $t_{end} = 3 \times 10^2, 10^3, 3 \times 10^3, 10^4$, and 3×10^4. (a) Entropy-based model. (b) Barabasi–Albert model. In both cases a new node is introduced every $\Delta t = 0.5$ and the node degrees evolve according to (26) or (3), in (a) and (b) respectively. Parameter values: $m = 5$, $\epsilon = 0.1$. For illustration we have plotted the power law distribution lines at $\gamma = -3$ and -2.5.

values 3×10^2, 10^3, 3×10^3, 10^4, and 3×10^4. The degree distributions for the entropy-based model do not clearly follow any power law behaviour, at least in the regime explored here, while the Barabasi–Albert model quickly assumes a form very close to a power-law degree distribution with exponent $\gamma = 3$ as we expect.

While any systematic analysis of (26) seems difficult, for large enough networks we might expect that this model is comparable to the classes of sub-linear preferential attachment models studied rigorously by Dereich & Mörters [39,40]. These authors prove that preferential attachment rules based on concave functions of node degree will asymptotically result in degree distributions with exponent $\gamma = 3$. This suggests that the long time dynamics of the entropy-based model might also show this behavior, but at intermediate times the more complex distributions illustrated in figure 6(a) might well be more typical.

5 Conclusion and Future Directions

In Section 2 we introduced a modification to the preferential attachment model to account for the maximum connections a node may have in a network. From the mathematical analysis we were able to predict both the value of the power law exponent γ and the presence of a hard limit on the degree distribution. In Section 3 we applied the analysis to an extensive range of social, citation and physical infrastructure graphs, and found that the constraint model's values for γ more accurately fitted the data. In addition, the constrained model implicitly contains a hard limit in the node degree, and the data analyzed had degree distributions with far fewer nodes of extremely large k than a pure power law would predict. This is an important result because the value is arrived at as a natural consequence of the presence of constraints on the maximum node degree, rather than by introducing a distribution of additional parameters such as in the fitness model.

Fitness is a valuable concept, and indeed in further work it is intended to investigate the role of a top constraint in a model extended to include the concept of fitness, or indeed generalized in a similar way to the NGF models. In particular the analogy with Bose-Einstein statistical mechanics is interesting, and opens up many applications of network science in more general theoretical physics, but the method outlined in this paper captures the essential features of real degree distributions without requiring the concept of fitness.

Motivated by the interesting results when applying concepts from statistical mechanics, and the results for vertex entropy arrived at in [8], we also set out to see if scale free models could be arrived at from pure thermodynamic principles of entropic force. In Section 4 we were able to obtain, from first principles, an evolution equation for the degree of a random node, which although soluble analytically, presents challenges when deriving the degree distributions according to the continuum analysis. The Taylor series for $log(x)$ converges only for values of x in the range $0 < x \leq 2$, but as $k \leq 2mt$, and, both terms are always strictly positive, we can safely expand the log term in equation (25). The validity of this expansion is not valid for $k \ll 2mt$ as the series for $log(x)$ converges very slowly as $x \to 0$. However at early times after the introduction of the node into the graph, $\frac{k}{2mt}$ will be closer to 1 and we can expand the log to yield:

$$\log\left(\frac{k}{2mt}\right) \approx \frac{k}{2mt} - 1 + \text{higher order terms.}$$

For the same period of time this expression is valid we can see that the leading terms in this expansion contribute to the ODE time evolution of k the following:

$$\frac{dk}{dt} \approx \frac{\epsilon k}{2t} - \frac{\epsilon k^2}{2mt} + \text{ higher order terms.} \qquad (27)$$

What can be asserted is that for a period of time after a node is introduced into the network its behavior will be governed by the first terms in this expansion, with much more complex behavior as the network evolves. This is illustrated nicely in Figure 6 obtained from our numerical simulations. These first two terms in the expansion are identical in form to the evolution of k with time in the Barabási-Albert model, *and also* a correction identical in form to our constrained model. This would indicate that for small t the behavior of the entropic model should closely resemble scale free, with a correction for constraints. As t increases the model will become more complex.

The model introduces ϵ as a free parameter, and it is a legitimate question to ask what the correct value of this should be. In the numerical simulations we chose, for illustrative purposes, $\epsilon = 0.1$. The choice of ϵ will have a profound affect on the family of graphs that can emerge from the initial conditions and in particular the slope of the power law degree distribution obtained. For example, values of $\epsilon > 1$ will tend to generate power laws with $\gamma < 3$, and conversely $\epsilon < 1$ will produce $\gamma > 3$, at least in the regime where the first term of equation(27) dominates. Given that the origin of the parameter is in the relative entropic force of the graph compared to a randomly picked node of degree k, one could speculate that its value measures the relative affect of an additional link on the bulk of the graph to increase entropy compared to an individual node of varying degree. High values of ϵ perhaps indicate relatively more homogeneous graphs than low values, indicating that degree distributions drop off more slowly the more ordered a graph's initial state. In future work we intend to investigate the dependency of graph evolution on ϵ in more detail, and whether the more complex evolution behavior of our dynamic model has utility in revealing more detail on the internal structure of dynamically evolving graphs.

We believe that there is a deep connection between vertex entropy and the evolution of networks. An attractive feature of our model is that it predicts scale free and more complex network evolution behavior from a first principles argument without appeal to any heuristics, node by node parameters, or indeed a stated but not justified property of nodes to seek out other high degree nodes with which to preferentially attach. Instead we argue from the safety of the second law of thermodynamics to a model which reproduces the essential features of scale freedom, and also the constrained model which we demonstrated provides a better fit to the experimental data. It is possible that higher terms in the expansion of equation (25) could yield insight into the detailed evolution of networks, and provide powerful analytical tools to for example determine the age of a network. Nevertheless, it is attractive to speculate that scale freedom, and similar models, may be a manifestation of the second law of thermodynamics as applied to graph evolution.

Beyond investigating the entropic model, there are many potential enhancements to the constrained model. In further work we intend to conduct analysis of more network datasets and also investigate corrections to the constrained model to improve our estimate of $(c - 2m)$ or $(c - \langle k \rangle)$ for the average occupancy of a node, by iterating the resultant distribution in equation (16) to calculate $\langle k \rangle$ as $\langle k \rangle = \int_{-\infty}^{+\infty} kP(k)dk$.

Author contribution statement

P.T., I.J.W. and G.P. developed the initial model and wrote the initial draft of the paper. I.Z.K. and P.T. refined the exposition of the constrained attachment and continuum models and resolved some technical issues. J.H.P.D. provided subsequent analysis and the results shown in Figure 6. All authors contributed to editing and production of the final manuscript.

A - Derivation of γ in Constrained Attachment

We recall from the main body of Section 2 our expression for $P(k)$ in Equation (16):

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

We can simplify this by collapsing the uninteresting details as follows:

$$P(k) = \frac{A(B-k)^{\frac{2}{\alpha}-1}}{k^{\frac{2}{\alpha}+1}}, \text{ where} \qquad (28)$$

$$A = \frac{2(c+\delta)\rho^{2/\alpha}t}{\alpha(t+m_0)}, \text{ and } B = (c+\delta)$$

Now, as $a^b = \exp\{b\log(a)\}$, we can write $(B-k)^{\frac{2}{\alpha}-1} = \exp\{(\frac{2}{\alpha}-1)\log(B-k)\}$. Substituting back into Equation (28), and taking the logarithm of both sides, we obtain:

$$\log(P(k)) = \log A + \Big(\frac{2}{\alpha}-1\Big)\log(B-k) - \Big(\frac{2}{\alpha}+1\Big)\log(k) \ .$$

We can further simplify by noting that $\log(B-k) = \log\{B(1-\frac{k}{B})\} = \log B + \log\Big(1-\frac{k}{B}\Big)$. We note that if $k \ll B$, either by taking small values of k or allowing $c \to \infty$, then $\frac{k}{B} \to 0$, so that $\log(B-k) = \log B + \log(1+0) = \log B$. Bringing this altogether we have:

$$\log(P(k)) = \log A + \Big(\frac{2}{\alpha}-1\Big)\log B - \Big(\frac{2}{\alpha}+1\Big)\log(k) \ .$$

Taking the exponential of both sides we end with the main result:

$$P(k) = \frac{2(c+\delta)\rho^{2/\alpha}t}{\alpha(t+m_0)} \times \frac{(c+\delta)^{(\frac{2}{\alpha}-1)}}{k^{(\frac{2}{\alpha}+1)}} \ ,$$
$$\text{which is of the form,} \qquad (29)$$
$$P(k) \propto \frac{1}{k^{(\frac{2}{\alpha}+1)}} \ .$$

In Equation (29), we arrive at the familar form of a scale free distribution with $\gamma = 2/\alpha + 1$. It is interesting to note that, as $c > 2m$, by definition, $\alpha \geq 1$ with equality in the limit that $c \to \infty$. This yields a range for the power law exponent γ as $1 \leq \gamma \leq 3$, with the familiar result of $\gamma = 3$ recovered in the case of the constraint being infinite, and therefore unimportant to the dynamics of the network growth.

We can also examine Equation (28) in the asymptotic limit of $c \to \infty$. We recall that $\rho = \frac{m}{c+\delta-m}$, and that $\alpha = \frac{c+\delta}{c-2m}$. At the limit $c \to \infty$, $\alpha = 1$, which reduces Equation (28) to:

$$P(k) \approx \frac{2c(\frac{m}{c})^2 t}{(t+m_0)} \times \left\{ \frac{c}{k^3} - \frac{1}{k^2} \right\} ,$$
$$\text{which multiplying out and allowing } c \to \infty, \text{ gives} \qquad (30)$$
$$P(k) \approx \frac{2m^2 t}{(t+m_0)} \times \frac{1}{k^3} \ .$$

As expected, this is precisely the form of the degree distribution in the standard preferential attachment model, which emerges as the constraint becomes infinite, and therefore unimportant in the dynamical growth of the network.

References

1. R. Albert, A.L. Barabási, Review of Modern Physics **74** (2002)
2. B. Bollobás, *Random Graphs*, 2nd edn. (Cambridge University Press, 2001), ISBN 9780521797221, http://dx.doi.org/10.1017/CBO9780511814068
3. Ã. Albert, H. Jeong, **272**, 173 (1999), 9907068v1
4. D. Watts, S. Strogatz, Nature **393**, 440 (1998), 0803.0939v1
5. S.N. Dorogovtsev, J.F.F. Mendes, A.N. Samukhin, Physical Review Letters **85**, 4633 (2000), 0004434
6. G. Bianconi, A.L. Barabási, Physical Review Letters **86**, 5632 (2001), 0011224
7. P.L. Krapivsky, S. Redner, F. Leyvraz, Physical Review Letters **85**, 4629 (2000), 0005139
8. P. Tee, G. Parisis, I. Wakeman, *Towards an Approximate Graph Entropy Measure for Identifying Incidents in Network Event Data*, in *IEEE/IFIP Network Operations and Management Symposium, NOMS* (Istanbul, Turkey, 2016), pp. 1049–1054, ISBN 9781509002238
9. P. Tee, G. Parisis, I. Wakeman, IEEE Transactions on Network and Service Management **PP**, 1 (2017)
10. M. Faloutsos, P. Faloutsos, C. Faloutsos, In SIGCOMM pp. 251–262 (1999)
11. R. Albert, H. Jeong, A. Barabasi, Nature **406**, 378 (2000), 0008064
12. P.R. Guimarães, M.A.M. De Aguiar, J. Bascompte, P. Jordano, S.F.D. Reis, Physical Review E - Statistical, Nonlinear, and Soft Matter Physics **71**, 3 (2005)
13. N. Berger, C. Borgs, J.T. Chayes, A. Saberi, Annals of Probability **42**, 1 (2014), 1401.2792
14. B.W. Herr, W. Ke, E. Hardy, K. Borner, Proceedings of the International Conference on Information Visualisation **2007**, 465 (2007)
15. R.M. D'Souza, C. Borgs, J.T. Chayes, N. Berger, R.D. Kleinberg, Proceedings of the National Academy of Sciences of the United States of America **104**, 6112 (2007)
16. G.B.a.L. Barabási, Europhysics Letters **54**, 13 (2000), 0011029
17. P. Moriano, J. Finke, Physical Review E - Statistical, Nonlinear, and Soft Matter Physics pp. 1090–1095 (2013)
18. G. Su, X. Zhang, Y. Zhang, EPL (Europhysics Letters) **100**, 38003 (2012), 1103.3196
19. J. Park, M.E.J. Newman, Physical Review E - Statistical, Nonlinear, and Soft Matter Physics **70**, 1 (2004), 0405566
20. G. Bianconi, C. Rahmede, Physical Review E - Statistical, Nonlinear, and Soft Matter Physics **93**, 1 (2016), 1511.04539
21. O.T. Courtney, G. Bianconi, Phys. Rev. E **95**, 062301 (2017)
22. R. Sharma, R. Bhandari, Communications in Statistics - Theory and Methods **43**, 4503 (2014)
23. A. Clauset, C. Rohilla Shalizi, M.E. J Newman, SIAM Review **51**, 661 (2009), arXiv:0706.1062v2
24. J. Leskovec, A. Krevl, *SNAP Datasets: Stanford Large Network Dataset Collection*, url{http://snap.stanford.edu/data} (2014)
25. M. Cha, H. Haddai, F. Benevenuto, K.P. Gummadi, International AAAI Conference on Weblogs and Social Media pp. 10–17 (2010)
26. S. Knight, H.X. Nguyen, N. Falkner, R. Bowden, M. Roughan, IEEE Journal on Selected Areas in Communications **29**, 1765 (2011)
27. J. Leskovec, J. Kleinberg, C. Faloutsos, ACM Transactions on Knowledge Discovery from Data **1**, 1 (2007), 0603229v3
28. L. Takac, M. Zabovsky, International Scientific Conference & International Workshop pp. 1–6 (2012)
29. J. Leskovec, J. Mcauley, Advances in neural information processing ... pp. 1–9 (2012), 1210.8182
30. J. Leskovec, K.J. Lang, A. Dasgupta, M.W. Mahoney, Internet Mathematics **6**, 29 (2011), 0810.1355
31. J. Patokallio, *Open Flights*, http://openflights.org/about (2016)
32. E. Schrödinger, *Statistical Thermodynamics* (Courier Corporation, 1989), ISBN 0486661016, 9780486661018
33. J. Körner, *FredmanKomlós bounds and information theory* (1986)
34. G. Simonyi, Combinatorial Optimization **20**, 399 (1995)
35. F. Passerini, S. Severini, ArXiv e-prints p. 5 (2008), 0812.2597
36. G. Bianconi, EPL (Europhysics Letters) **81**, 28005 (2008), 0708.0153
37. M. Dehmer, Applied Mathematics and Computation **201**, 82 (2008)
38. M. Dehmer, A. Mowshowitz, Information Sciences **181**, 57 (2011)

39. S. Dereich, P. Mörters, Electronic Journal of Probability **14**, 1222 (2009)
40. S. Dereich, P. Mörters, Jahresaber. Dtsch. Math-Ver. **113**, 21 (2011)

6.2 Discussion

The paper presented in this chapter contains two specific contributions, the first a concrete model of network evolution in the presence of degree constraints, and the second a speculative investigation of the extent to which considerations of entropy underpin the mechanisms of random network growth. This second part of the paper is worthy of some additional discussion, and indeed in the review process for the paper, the far reaching implication of the second law being the actual mechanism behind dynamic network growth was singled out as distracting from the main discourse of the constrained attachment model. We subsequently changed the title of the paper from "Is Preferential Attachment the 2^{nd} law of Thermodynamics in Disguise" to "Constraints and Entropy in a Model of Network Evolution" to respond to this comment. Nevertheless, I stand behind the belief that the maximization of informational entropy is a valid constraint in random network evolution processes, and I lay out here some of the supporting arguments.

In the general literature of theoretical physics there is increased interest in the *'it from bit'* hypothesis of nature [64]. This has included a new approach to gravity as an emergent force developed by Verlinde *et al* [79, 80, 35], an attempt to explain Brownian motion using the same entropic approach by Roos [60], and even attempts to describe the electrostatic 'Coulomb' force as entropic in origin [81]. It is this broad consideration of entropy as the fundamental principle behind a wide range of physical phenomena that motivated the approach described in Section 4 of the paper in this Chapter.

In each of these treatments the starting point is entropic force. This is a well understood property of physical processes that have an element of equilibrium choice, which when treated using a stochastic process exhibit a measurable force that drives the system towards maximum entropy. The classic example of this is the contraction of a polymerized molecule in a heat bath, which is exhibited when the polymer has a series of molecular 'joints' that can be freely oriented in space. The system will 'choose' a configuration of molecular angles such that entropy is maximized, and this generally leads to a more compact molecule. If the molecule is stretched out, this forces the orientation of the molecular bonds into a more consistent and therefore ordered and lower entropy state, and it resists the pull with a force $\mathbf{F}$, equal to $\mathbf{F} = \mathbf{T}\Delta\mathbf{S}$. Similar arguments exist for the cellular process of Osmosis and are surveyed in [80, 60].

In the case of graph evolution, the paper attempts to set out the sketch of an approach based around the maximization of graph entropy as the driver of network growth. In the set up of the argument, you consider the classic case of the preferential attachment model that

envisages the addition of a node at unit time step that randomly connects to m other nodes. Crucially, rather than asserting that the probability of selecting a given node is determined by its relative degree, instead we consider an ensemble of many different graphs that share the same node and edge count of the graph at this point in its evolution. We can then assert that the most likely nodes selected will be precisely those that maximize the graph's entropy.

To quantify this selection probability, we use the vertex entropy introduced in the paper presented in Chapter 5, and assert that the probability of a node being selected for attachment is proportional to the relative entropic force the new node has to connect to the selected node, versus all other nodes. Now it should be emphasized that I am not suggesting that this force is a real measurable force. Indeed from dimensional analysis all informational entropy is dimensionless, and would need to be multiplied by a constant (in statistical physics this is the famous Boltzmann constant), to deliver a physical force. However the argument is sound as it is a comparative 'force' that factors out inconvenient physical constants and quantities such as temperature.

The analysis in the paper, via an approximation of the vertex entropy of a randomly chosen graph in the ensemble, yields the following result for the time evolution of a node's degree:

$$\frac{dk_i}{dt} = 2m\Pi_i = -\varepsilon\frac{k_i}{t}\left\{\frac{1}{2} + \log\left(\frac{k_i}{2mt}\right)\right\}. \tag{6.1}$$

The central thrust of the argument is that this form of degree evolution replicates many of the features of scale free network, including both linear and logarithmic cut-off terms in the differential equation. It is clear that this does not amount to a proof of the link between vertex entropy and scale free network evolution, but it as an avenue for future investigation to explore this link further.

As a direction of research the approach is not isolated. For example in the paper by Newman *et al* [56], the approach of analyzing the behavior of dynamic networks using a statistical mechanical analysis of network ensembles is extensively explored. In this work, the analysis considers ensembles of graphs $\mathscr{G}$, in which a particular graph $G \in \mathscr{G}$ occurs with a probability $P(G)$. This probability is constrained to sum to unity, and generate some measurable graph property as an expectation value when summed across all graphs. For example let $m(G)$ be the number of edges of a graph, the constraint would read $\sum_{G\in\mathscr{G}} P(G)m(G) = \langle m\rangle$, and $\sum_{G\in\mathscr{G}} P(G) = 1$. Entropy of this ensemble is then simply defined as $S = -\sum_{G\in\mathscr{G}} P(G)log_2P(G)$. The analysis proceeds in direct analogy with statistical physics,

defining a partition function Z (for a classic treatment of partition functions in statistical physics see [61]), which describes the precise distribution of the ensemble across possible configurations or states, and an implicit maximization of entropy. Newman refers to these graphs as 'exponential random graphs'.

This approach has been extended by many authors, notably Bianconi *et al* in [10, 3] where the analysis is further constrained to consider only ensembles of networks with identical degree distributions. In earlier work by Bianconi *et al*, a series of articles develops this analogy with statistical physics, and leads to the prediction of Bose-Einstein condensation of random networks [7, 11]. The analogy between random graphs and statistical thermodynamics is well established, and has led to deeper insight into the dynamics of network evolution.

In this context the proposal in the work presented in this Chapter can be contextualized. Fundamentally the contribution avoids working backwards from an unknown probability distribution of graph configurations (the precise arrangement of $|V|$ vertices and $|E|$ edges in a graph $G(V,E)$), in an ensemble to measures of entropy and a partition function. Instead our analysis starts with an estimation of the vertex entropy for a randomly selected node, in a random member of a graph ensemble, at a fixed point in the evolution of all of the graphs in the ensemble. Using this estimation we can derive Equation (6.1) to describe the degree evolution dynamics of *any* node in the ensemble of graphs. It is encouraging for further work that even with such a broad estimation of vertex entropy, it is possible to infer linear and non-linear node degree evolution.

Chapter 7

Conclusion and Future Directions

The papers presented in Chapters 4,5 and 6 form a continuous thread investigating the possibilities and implications of defining a measure of structural graph entropy at the node or vertex level of a graph. In the first two papers this is specifically confined to an important problem in the operational management of communications networks, and I was able to prove that the measures are practically useful in identifying important nodes from a monitoring perspective.

The final paper may appear to be a disconnected problem, but the principle motivation for analyzing dynamic network growth was to identify whether vertex entropy could be of use in explaining deviations of real network structure from the current best theory preferential attachment. The paper successfully concludes that a dynamic theory using vertex entropy could explain preferential attachment and also a basic extension that was proposed to better model networks that have a natural degree constraint. In section 5.2.1 of the thesis, an analysis is conducted to compare the values of our vertex entropy measures, summed across the whole graph, with global entropy measures. The conclusion of this analysis is that there could be a relationship between our proposed form of vertex entropy and the more established, global, formulations of entropy.

During the course of the research both of the major avenues of investigation have interesting open questions.

- **Vertex Entropy :** The analysis of vertex entropy relied upon the definition of a j-Sphere, which we exclusively used for values of $j = 1$. No investigation has been undertaken at values of $j > 1$, and this is an important open question. There are also many other alternative definitions of node importance such as betweenness and eigenvalue centrality, and it is interesting to ask how vertex entropy and centrality may

be related. In the case of eigenvalue centrality, the definition may well have a profound relationship to Von Neumann entropy. Only limited analysis was undertaken to experimentally verify whether the local vertex entropy, summed across the whole graph, is correlated with either structural or chromatic entropy. However, the results offered some encouragement that such a correlation exists, which is an exciting possibility as this would open up the possibility of a computationally cheap way to approximate the calculation of graph entropy, a known NP-Hard problem.

- **Constrained Attachment :** There are a number of assumptions made in the constrained attachment model that could be modified and investigated. In particular. it is assumed that both $\langle c_i(t) \rangle$, and δ are constants (the latter we explicitly prove in simulations), but it would be interesting to extend the model to allow for these quantities to vary with time. The model also only admits one capacity limitation, but in real networks this would actually be multiple constraints, which may themselves change over time or be subject to a distribution. Extending the model to admit these changes may lead to even better fit to the experimental data.

  Regarding the entropic model of network growth, only one of the variants of vertex entropy was used to construct the model, and an obvious extension would be to consider other definitions. The solution of the resultant equation for $P(k)$, was only possible numerically, and it may be that the problem is amenable to a perturbation style approximation, which may yield interesting insight into higher order corrections to preferential attachment. It is also noted that the additional structure in the model may make it possible to identify how mature a network is in its dynamic evolution, which could lead to many practical applications.

It is the case that both research topics have a rich set of questions to further investigate, and beyond this work I look forward to engaging with them!

References

[1] R Albert, H Jeong, and AL Barabasi. Error and attack tolerance of complex networks. *Nature*, 406(6794):378–82, 2000. ISSN 1476-4687. doi: 10.1038/35019019.

[2] Reká Albert and Albert-László Barabási. Statistical mechanics of complex networks. *Review of Modern Physics*, 74(January), 2002.

[3] Kartik Anand, Ginestra Bianconi, and Simone Severini. Shannon and von Neumann entropy of random networks with heterogeneous expected degree. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 83(3):1–10, 2011. ISSN 15393755. doi: 10.1103/PhysRevE.83.036109.

[4] Ravindra B. Bapat. *Graphs and Matrices*. Springer-Verlag London, 2010. doi: 10.1007/978-1-84882-981-7.

[5] A. L. Barabási, Réka Albert, and Hawoong Jeong. Mean-field theory for scale-free random networks. *Physica A: Statistical Mechanics and its Applications*, 272(1): 173–187, 1999. ISSN 03784371. doi: 10.1016/S0378-4371(99)00291-5.

[6] Albert-Lszl Barabási. *Network Science*. Cambridge University Press; 1 edition (August 5, 2016), 2016. ISBN 978-1107076266.

[7] G. Bianconi a. L. Barabási. Competition and multiscaling in evolving networks. *Europhysics Letters*, 54(May):13, 2000. ISSN 0295-5075. doi: 10.1209/epl/i2001-00260-6.

[8] Leila Bennacer, Laurent Ciavaglia, Samir Ghamri-Doudane, Abdelghani Chibani, Yacine Amirat, and Abdelhamid Mellouk. Scalable and fast root cause analysis using inter cluster inference. *IEEE International Conference on Communications*, pages 3563–3568, 2013. ISSN 15503607. doi: 10.1109/ICC.2013.6655104.

[9] Noam Berger, Christian Borgs, Jennifer T. Chayes, and Amin Saberi. Asymptotic behavior and distributional limits of preferential attachment graphs. *Annals of Probability*, 42(1):1–40, 2014. ISSN 00911798. doi: 10.1214/12-AOP755.

[10] Ginestra Bianconi. The entropy of randomized network ensembles. *EPL (Europhysics Letters)*, 81(2):28005, 2008. ISSN 0295-5075. doi: 10.1209/0295-5075/81/28005.

[11] Ginestra Bianconi and Albert-László Barabási. Bose-Einstein Condensation in Complex Networks. *Physical Review Letters*, 86(24):5632–5635, 2001. ISSN 0031-9007. doi: 10.1103/PhysRevLett.86.5632.

[12] Ginestra Bianconi and Christoph Rahmede. Network geometry with flavor: From complexity to quantum geometry. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 93(3):1–37, 2016. ISSN 15502376. doi: 10.1103/PhysRevE.93.032315.

[13] Béla Bollobàs. *Modern Graph Theory*. Springer-Verlag New York, 1998. doi: 10.1007/978-1-4612-0619-4.

[14] Béla Bollobás. *Random Graphs*. Cambridge University Press, 2nd edition, 2001. ISBN 9780521797221.

[15] C. Boutsidis and E. Gallopoulos. SVD based initialization: A head start for nonnegative matrix factorization. *Pattern Recognition*, 41(4):1350–1362, apr 2008. ISSN 00313203. doi: 10.1016/j.patcog.2007.09.010.

[16] AP Andrew P. Bradley. The use of the area under the ROC curve in the evaluation of machine learning algorithms, 1997. ISSN 00313203. URL http://www.sciencedirect.com/science/article/pii/S0031320396001422.

[17] Alfonso Castro, V Villagra, Beatriz Fuentes, and Begoña Costales. A flexible architecture for service management in the cloud. 11(1):116–125, 2014. doi: 10.1109/TNSM.2014.022614.1300421.

[18] Census, US. E-Stats 2014: Measuring the Electronic Economy, 2016. URL http://www.census.gov/content/dam/Census/library/publications/2016/econ/e14-estats.pdf.

[19] Aaron Clauset, Cosma Rohilla Shalizi, and M E J Newman. Power-Law Distributions in Empirical Data. *SIAM Review*, 51(4):661–703, 2009. ISSN 19417330. doi: 10.1214/13-AOAS710.

[20] Gregory F Cooper. The computational complexity of probabilistic inference using bayesian belief networks. *Artificial Intelligence*, 42(2-3):393–405, 1990. ISSN 00043702. doi: 10.1016/0004-3702(90)90060-D.

[21] Owen T. Courtney and Ginestra Bianconi. Weighted Growing Simplicial Complexes. pages 1–18, 2017. ISSN 2470-0045. doi: 10.1103/PhysRevE.95.062301. URL http://arxiv.org/abs/1703.01187.

[22] Imre Csiszár. Axiomatic characterizations of information measures. *Entropy*, 10(3):261–273, 2008. ISSN 10994300. doi: 10.3390/e10030261.

[23] Matthias Dehmer. Information processing in complex networks: Graph entropy and information functionals. *Applied Mathematics and Computation*, 201(1-2):82–94, 2008. ISSN 00963003. doi: 10.1016/j.amc.2007.12.010.

[24] Matthias Dehmer and Abbe Mowshowitz. A history of graph entropy measures. *Information Sciences*, 181:57–78, 2011. ISSN 00200255. doi: 10.1016/j.ins.2010.08.041.

[25] DellEMC Inc. EMC Automated Data Center Manager, 2017. URL http://www.emc.uz/it-management/smarts/index.htm.

[26] Debora Donato, Luigi Laura, Stefano Leonardi, and Stefano Millozzi. The Web as a graph. *ACM Transactions on Internet Technology*, 7(1):4–es, 2007. ISSN 15335399. doi: 10.1145/1189740.1189744.

[27] S. N. Dorogovtsev, J. F F Mendes, and A. N. Samukhin. Structure of growing networks with preferential linking. *Physical Review Letters*, 85(21):4633–4636, 2000. ISSN 00319007. doi: 10.1103/PhysRevLett.85.4633.

[28] Raissa M D'Souza, Christian Borgs, Jennifer T Chayes, Noam Berger, and Robert D Kleinberg. Emergence of tempered preferential attachment from optimization. *Proceedings of the National Academy of Sciences of the United States of America*, 104(15): 6112–7, 2007. ISSN 0027-8424. doi: 10.1073/pnas.0606779104.

[29] Michalis Faloutsos, Petros Faloutsos, and Christos Faloutsos. On Power-Law Relationships of the Internet Topology. *In SIGCOMM*, pages 251–262, 1999. ISSN 01464833. doi: 10.1.1.37.234.

[30] Andy Field. *Discovering Statistics Using IBM SPSS Statistics*. Sage Publications Ltd., 4th edition, 2013. ISBN 1446249182,9781446249185.

[31] Maksym Gabielkov and Arnaud Legout. The Complete Picture Of the Twitter Social Graph. *CoNEXT*, pages 20–21, 2012. doi: 10.1145/2413247.2413260.

[32] R.D Gardner and D.a. Harle. Methods and systems for alarm correlation. In *Proceedings of GLOBECOM'96. 1996 IEEE Global Telecommunications Conference*, volume 1, pages 136–140, London, 1996. ISBN 0-7803-3336-5. doi: 10.1109/GLOCOM.1996.594348. URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=594348.

[33] GraphML Team. Graphml file format for graphs., 2002. URL http://graphml.graphdrawing.org/.

[34] Petter Holme, Josh Karlin, and Stephanie Forrest. An integrated model of traffic, geography and economy in the internet. *ACM SIGCOMM Computer Communication Review*, 38(3):5, 2008. ISSN 01464833. doi: 10.1145/1384609.1384611.

[35] Sabine Hossenfelder. A Covariant Version of Verlinde's Emergent Gravity. 2017. URL https://arxiv.org/pdf/1703.01415.pdf.

[36] IBM Inc. IBM Tivoli Netcool/OMNIbus, 2017. URL http://www-03.ibm.com/software/products/en/ibmtivolinetcoolomnibus.

[37] ITU-T. STRUCTURE OF MANAGEMENT INFORMATION: GUIDELINES FOR THE DEFINITION OF MANAGED OBJECTS. Technical report, ITU-T, 1992.

[38] N P Khomenko and L D Golovko. Identifying Certain Types of Parts of a Graph and Computing their Number. *Ukrainian Mathematical Journal*, 24(3):313–321, 1972.

[39] S Kliger, S Yemini, and Y Yemini. A coding approach to event correlation. ... *Network Management IV*, 1995. doi: 10.1007/978-0-387-34890-2_24.

[40] Simon Knight, Hung X. Nguyen, Nickolas Falkner, Rhys Bowden, and Matthew Roughan. The internet topology zoo. *IEEE Journal on Selected Areas in Communications*, 29(9):1765–1775, 2011. ISSN 07338716. doi: 10.1109/JSAC.2011.111002.

[41] János Körner. Fredman–Komlós bounds and information theory, 1986. ISSN 0196-5212.

[42] P. L. Krapivsky, S. Redner, and F. Leyvraz. Connectivity of growing random networks. *Physical Review Letters*, 85(21):4629–4632, 2000. ISSN 00319007. doi: 10.1103/PhysRevLett.85.4629.

[43] Jure Leskovec and Andrej Krevl. SNAP Datasets: Stanford Large Network Dataset Collection. url{http://snap.stanford.edu/data}, jun 2014.

[44] Jure Leskovec and Jj Mcauley. Learning to discover social circles in ego networks. *Advances in neural information processing ...*, pages 1–9, 2012. ISSN 15564681. doi: 10.1145/0000000.0000000.

[45] Jure Leskovec, Jon Kleinberg, and Christos Faloutsos. Graphs over Time : Densification Laws , Shrinking Diameters and Possible Explanations. 2005.

[46] Jure Leskovec, Jon Kleinberg, and Christos Faloutsos. Graph Evolution: Densification and Shrinking Diameters. *ACM Transactions on Knowledge Discovery from Data*, 1(2): 1–39, 2007. ISSN 15564681. doi: 10.1145/1217299.1217301.

[47] R.M.R. Lewis. A Guide to Graph Colouring. pages 1–253, 2016. doi: 10.1007/978-3-319-25730-3.

[48] Moogsoft Inc. Incident.MOOG Documentation 5.2.3, 2016. URL http://docs.moogsoft.com/display/050203/Incident.MOOG.

[49] Abbe Mowshowitz and Matthias Dehmer. Entropy and the complexity of graphs revisited. *Entropy*, 14:559–570, 2012. ISSN 10994300. doi: 10.3390/e14030559.

[50] Abbe Mowshowitz and Valia Mitsou. Entropy, Orbits, and Spectra of Graphs. *Analysis of Complex Networks: From Biology to Linguistics*, pages 1–22, 2009. doi: 10.1002/9783527627981.ch1.

[51] MycomOSI Inc. NetExpert Datasheet, 2016. URL http://www.mycom-osi.com/products/netexpert-fault-service-impact-management.

[52] Office of Government Commerce. *Service Operation Book*. The Stationery Office, 2007. ISBN 0113310463. URL https://www.itil.org.uk/so.htm.

[53] Adam Oliner and J Stearley. What supercomputers say: A study of five system logs. *Dependable Systems and Networks, 2007 ...*, 2007. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4273008.

[54] AJ Oliner, Alex Aiken, and Jon Stearley. Alert detection in system logs. *Data Mining, 2008. ICDM'08. ...*, 2008.

[55] Fragkiskos Papadopoulos, Maksim Kitsak, M. Angeles Serrano, Marian Boguna, and Dmitri Krioukov. Popularity versus Similarity in Growing Networks. pages 8–11, 2011. ISSN 0028-0836. doi: 10.1038/nature11459. URL http://arxiv.org/abs/1106.0286{%}0Ahttp://dx.doi.org/10.1038/nature11459.

[56] Juyong Park and M. E J Newman. Statistical mechanics of networks. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 70(6 2):1–13, 2004. ISSN 15393755. doi: 10.1103/PhysRevE.70.066117.

[57] Filippo Passerini and Simone Severini. The von Neumann entropy of networks. *ArXiv e-prints*, (12538):5, 2008. doi: 10.4018/978-1-60960-171-3.ch005. URL http://arxiv.org/abs/0812.2597.

[58] Jano Patokallio. Open Flights. http://openflights.org/about, 2016.

[59] D M W Powers. Evaluation: From Precision, Recall and F-Measure To Roc, Informedness, Markedness {&} Correlation. *Journal of Machine Learning Technologies*, 2(1): 37–63, 2011. ISSN 2229-3981. doi: 10.1.1.214.9232.

[60] Nico Roos. Entropic forces in Brownian motion. *Entropic forces in Brownian motion*, (1):1–10, 2013. ISSN 19432909. doi: 10.1119/1.4894381. URL http://arxiv.org/abs/1310.4139.

[61] Erwin Schrödinger. *Statistical Thermodynamics*. Courier Corporation, 1989. ISBN 0486661016, 9780486661018.

[62] Claude Elwood Shannon. A Mathematical Theory of Communication. *The Bell System Technical Journal*, 27:379–423, 1948. ISSN 0724-6811.

[63] Gábor Simonyi. Graph entropy: a survey. *Combinatorial Optimization*, 20:399–441, 1995.

[64] Robert Spekkens. The invasion of physics by information theory. 1:1–42, 2016. ISSN 1098-6596. doi: 10.1017/CBO9781107415324.004.

[65] Leo Spizzirri. Justification and Application of Eigenvector Centrality. *Math.Washington.Edu*, 2011. URL https://www.math.washington.edu/~morrow/336_11/papers/leo.pdf.

[66] Jon Stearley. Towards informatic analysis of syslogs. ... *Computing, 2004 IEEE International Conference on*, 2004.

[67] Jon Stearley. Root cause analysis. Technical report, Sandia Labs, 2006. URL http://www.csm.ornl.gov/srt/conferences/ResilienceSummit/2008/speakers/stearley.html.

[68] Ma Lgorzata Steinder and Adarshpal S. Sethi. A survey of fault localization techniques in computer networks. *Science of Computer Programming*, 53(2):165–194, nov 2004. ISSN 01676423. doi: 10.1016/j.scico.2004.01.010.

[69] James V Stone. *Information Theory: A Tutorial Introduction*. Sebtel Press, 2013. ISBN 0956372856, 9780956372857.

[70] Lubos Takac and M Zabovsky. Data Analysis in Public Social Networks. *International Scientific Conference & International Workshop*, (May):1–6, 2012. URL http://snap.stanford.edu/data/soc-pokec.pdf.

[71] Yongming Tang, E Al-Shaer, and Raouf Boutaba. Efficient fault diagnosis using incremental alarm correlation and active investigation for internet and overlay networks. *Network and Service ...*, 5(1):36–49, 2008.

[72] P Tee. The role of graph entropy and constraints on fault localization and structure evolution in data networks. In *Mathematics of Networking 15, http://www.monmeetings.org/meeting15*, 2016. URL http://www.monmeetings.org/meeting15/.

[73] P Tee, I Wakeman, G Parisis, J Dawes, and I Kiss. Constraints and Entropy in a Model of Network Evolution. *ArXiv e-prints*, dec 2016. doi: 10.1140/epjb/e2017-80185-5. URL https://arxiv.org/abs/1612.03115v3.

[74] P. Tee, G. Parisis, and I. Wakeman. Vertex entropy as a critical node measure in network monitoring. *IEEE Transactions on Network and Service Management*, PP(99):1–1, 2017. ISSN 1932-4537. doi: 10.1109/TNSM.2017.2724301.

[75] Phil Tee, George Parisis, and Ian Wakeman. Towards an Approximate Graph Entropy Measure for Identifying Incidents in Network Event Data. (AnNet):1049–1054, 2016. doi: 10.1109/NOMS.2016.7502959.

[76] Tian Bu and D. Towsley. On distinguishing between Internet power law topology generators. *Proceedings.Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies*, 2(c):638–647, 2002. ISSN 0743-166X. doi: 10.1109/INFCOM.2002.1019309. URL http://ieeexplore.ieee.org/document/1019309/.

[77] University of Nebraska. Using the Receiver Operating Characteristic (ROC) curve to analyze a classification model, 1993. URL http://www.math.utah.edu/~gamez/files/ROC-Curves.pdf.

[78] P. Varga and L. Moldovan. Integration of service-level monitoring with fault management for end-to-end multi-provider ethernet services. *IEEE Transactions on Network and Service Management*, 4(1):28–38, jun 2007. ISSN 1932-4537. doi: 10.1109/TNSM.2007.030103.

[79] Erik Verlinde. On the Origin of Gravity and the Laws of Newton . *Journal of High Energy Physics*, 2011. doi: 10.1007/JHEP04(2011)029.

[80] Erik P. Verlinde. Emergent Gravity and the Dark Universe. *arXiv*, pages 0–50, 2016. URL http://arxiv.org/abs/1611.02269.

[81] Tower Wang. Coulomb force as an entropic force. *Physical Review D - Particles, Fields, Gravitation and Cosmology*, 81(10), 2010. ISSN 15507998. doi: 10.1103/PhysRevD.81.104045.

[82] Scott White and Padhraic Smyth. Algorithms for estimating relative importance in networks. *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '03*, page 266, 2003. doi: 10.1145/956755.956782.

[83] H. S Wilf and G Szekeres. An Inequality for the Chromatic Number of a Graph. *Journal of Combinatorial Theory*, 4(1):1–3, 1968. ISSN 00045411. doi: 10.1016/S0021-9800(68)80081-X.

[84] Walter Willinger, David Alderson, and John C Doyle. Mathematics and the Internet: A Source of Enormous Confusion and Great Potential. *Notices of the AMS*, 56(5):586–599, 2009. ISSN 00029920. doi: 10.1.1.152.2425.

[85] DCM Wood, SS Coleman, and MF Schwartz. Fremont: A System for Discovering Network Characteristics and Problems. *USENIX Winter*, pages 335–348, 1993. URL http://dsc.ufcg.edu.br/~jacques/cursos/2001.2/projii/Autotopologia/recursos/Fremont.pdf.

[86] Xindong Wu, Vipin Kumar, Quinlan J. Ross, Joydeep Ghosh, Qiang Yang, Hiroshi Motoda, Geoffrey J. McLachlan, Angus Ng, Bing Liu, Philip S. Yu, Zhi Hua Zhou, Michael Steinbach, David J. Hand, and Dan Steinberg. *Top 10 algorithms in data mining*, volume 14. 2008. ISBN 1011500701. doi: 10.1007/s10115-007-0114-2.

[87] Cheng Zhang, Jianxin Liao, and Xiaomin Zhu. SWPM: An Incremental Fault Localization Algorithm Based on Sliding Window with Preprocessing Mechanism. *2008 Ninth International Conference on Parallel and Distributed Computing, Applications and Technologies*, pages 235–242, 2008. doi: 10.1109/PDCAT.2008.57.

[88] Shi Zhou and Mondragón. Accurately modeling the internet topology. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 70(6 2):1–8, 2004. ISSN 15393755. doi: 10.1103/PhysRevE.70.066108.

[89] Hubert Zimmermann. Open Systems Interconnection. *IEEE Transactions on Communications*, 28(4):425–432, 1980. ISSN 00962244. doi: 10.1109/TCOM.1980.1094702.

Reference [75] first appeared in July 2016 IEEE proceedings of the NOMS 2016 Conference. Reprinted, with permission, from Phil Tee, George Parisis and Ian Wakeman, "Towards an approximate graph entropy measure for identifying incidents in network event data", Network Operations and Management Symposium (NOMS), 2016 IEEE/IFIP, published July 2016. Reference [74] will appear in 2017 IEEE Transactions on Network and Service Management. Reprinted, with permission, from Phil Tee, George Parisis and Ian Wakeman, "Vertex entropy as a critical node measure in network monitoring". Reference [73] will appear in 2017 European Physical Journal B. Reprinted, with permission, from Phil Tee, George Parisis, Ian Wakeman, Istv'an Kiss and Jonathan Dawes "Constraints and Entropy in a Model of Network Evolution".

Appendix A

Personal Biography

In my professional career, I have been involved in a considerable number of companies that have produced fault management products.My usual role in these businesses has been part of the founding team with particular responsibility for the invention and productization of the company's software products. This activity has spanned two and a half decades and the products built in these companies have been widely deployed in many large scale enterprises. This experience formed a strong part of the motivation for me to engage in my doctoral studies to deepen the theoretical understanding of the operation and limitations of these products.

A brief summary of these activities is as follows:

- **Avante Garde Computing:** I was involved in the engineering team, based in Sunnyvale, responsible for Net/Command. This product was a very early rules based management system, as described in Section 1.2.2, that associated every incoming alert with a script, written in REXX, that would perform diagnostic checks. I was involved in the engineering of a number of components, including the rules execution engine.

- **Micromuse Inc:** I was part of the founding team of Micromuse and was the primary inventor and designer of the company's product Netcool/OMNIbus. This product was the first software product to include an active in memory database that was capable of executing boolean logic on alerts with far greater throughput than competitive products, and as such is an example of a rules based system described in Section 1.2.2. It is still in use today in over 1000 large scale enterprises, including most communications service providers and banks. The company was listed on NASDAQ in 1998 and was acquired by IBM in 2006.

- **RiverSoft Plc:** I founded RiverSoft in 1998 and designed and built the company's product OpenRiver. The product was the first to use an 'Active Object' approach to modeling a network, and included a novel distributed network discovery engine as described in Section 1.2.2. The product has been deployed at over 1000 enterprises and is still in wide usage today. The company was listed on the London Stock Exchange in 2000 and was acquired by Micromuse in 2002.

- **Promethyan Labs LLC:** As part of the founding team of Promethyan I prototyped and designed the products that went on to be the core offering of a number of companies. Notable amongst those was Prelert Inc, which was acquired in 2017 by Elastic Search. It was an early example of a a data driven management application described in Section 1.2.2, and is now part of the Elastic Search suite of data management tools.

- **Moogsoft Inc:** I currently serve as Chief Executive, and Chief Technologist of Moogsoft Inc, the company I founded in 2012. At Moogsoft I designed the first fault management product that uses machine learning as the primary method to perform fault localization, and is an example of the approach described in Section 1.2.2. This has resulted in 15 patents being filed, with 5 in full grant. Today the product is in use at over 75 large scale operations and the company is expanding rapidly.