Change Point Detection in Social Networks –
– Critical Review with Experiments

Lucy Kendrick

Data Science Institute,Faculty of Science and Technology, Bournemouth University, BH12 5BB, Bournemouth, UK, i7949850@bournemouth.ac.uk

Katarzyna Musial and Bogdan Gabrys[1]

Advanced Analytics Institute, Faculty of Engineering and IT, University of Technology Sydney, Australia, {katarzyna.musial-gabrys, bogdan.gabrys}@uts.edu.au

Abstract

Change point detection in social networks is an important element in developing the understanding of dynamic systems. This complex and growing area of research has no clear guidelines on what methods to use or in which circumstances. This paper critically discusses several possible network metrics to be used for a change point detection problem and conducts an experimental, comparative analysis using the Enron and MIT networks. Bayesian change point detection analysis is conducted on different global graph metrics (Size, Density, Average Clustering Coefficient, Average Shortest Path) as well as metrics derived from the Hierarchical and Block models (Entropy, Edge Probability, No. of Communities, Hierarchy Level Membership). The results produced the posterior probability of a change point at weekly time intervals that were analysed against ground truth change points using precision and recall measures. Results suggest that computationally heavy generative models offer only slightly better results compared to some of the global graph metrics. The simplest metrics used in the experiments, i.e. nodes and links numbers, are the recommended choice for detecting overall structural changes.

[1]This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

[2]Katarzyna Musial is the corresponding author.

Keywords: Change Point Detection, Social Network, Dynamic Network

1. Introduction

For many years the analysis of complex networks remained a static exercise. Now research is increasingly viewing networks as dynamic systems, where the dynamic properties are as important as overall network structure. The computational capability to study not only large graphs, but a long sequence of large graphs over time has led to growing research in the field of detecting, modelling and predicting changes in complex networks [1, 2, 3, 4, 5, 6, 7]. The focus of this paper is on the problem of change point detection, which is a form of dynamic anomaly detection that has a long history of study in traditional time series datasets [8, 9, 10, 11, 12, 13].

There are many detection algorithms to find individual anomalies in static graphs [2]. These focus on the more traditional form of an anomaly that involves finding one unusual data point or node. The motivation behind this paper stems from the growing field of research that uses generative models to study change point detection in dynamic networks [14, 15, 3, 4, 6, 1]. Generative models are ways to probabilistically represent network data into sets of communities or hierarchy. It offers a potentially rich representation that can monitor smaller or subtle changes happening in sub-sections of a graph.

As a new area of research there is a need to establish the best ways to model the change point detection problem. There is also a lack of understanding in the generative model space of why one type of model should be selected over another. The aim of our research has therefore been to critically review the existing approaches and conduct an experimental analysis exploring different potential network metrics that can be used to detect changes in such complex, dynamic networks.

The paper begins with a review of the related work in Section 2 that provides a discussion on change point detection and the use of generative models in this research area. This is followed by Section 3 describing the metrics used in the experimental analysis. The datasets, experimental set up, the results of conducted experiments and the related discussions are presented in Section 4. Finally, Section 5 provides the conclusions and highlights some identified future research directions.

2. Related Work

The problem of Change Point Detection (CPD) historically stems from research assessing classical time series data to identify a change in the underlying mean or distribution of a given variable. Changes can be identified from calculations that measure the posterior probability of a change in monitored parameters. Such techniques have been successfully applied to many engineering and control problems to identify faults in systems [8, 13]. The overriding aim for CPD research, in the field of complex networks, is to identify a point in time where the graph exhibits a difference in behaviour. This time point can then be analysed to uncover an underlying cause.

Change Point Detection in complex networks is often tied to the field of anomaly detection. Both research areas use similar methods that exploit the existence of communities in graphs to establish unusual behaviour [2]. As a relatively new area of research there is no leading methodology used to conduct CPD in networks. According to a common methodology for CPD using time series analysis, the first step should be a preliminary investigation of the best way to model the problem followed by a selection of the best variables to be used as change indicators [8].

From the literature we find that change point and anomaly detection research will often use generative network models as a way to model the problem on a complex network. Generative models provide a well-recognised way of finding community structures or hierarchy in a graph with the additional benefit of using probabilistic values. Though most CPD studies agree on the use of generative models in this research area, they do not agree on any specific one to be clearly better than the others.

2.1. The Change Point Detection Problem

In the context of statistical methods employed, Basseville et al. [8] define three main problem areas in CPD:

– **On-line-detection**, where it is required that the change be identified as soon as possible to near real time. In the context of control problems this is often the main aim. This would ensure any faults in a monitored system caused by an unforeseen change can be highlighted instantly. This method, however, suffers from the issue of false alarms (false positives) where what may appear to be a change was only an anomaly.

– **Off-line hypothesis testing**, where the aim is to maximise the trade off between correctly identified change points and false alarms. This is often used as a retrospective analysis. This method has been often used as evidenced in the literature reviewed in the following sections.

– **Detecting the exact time of a change**, which can be used in combination with the above two approaches but where only one change point is to be discovered and it is assumed that no other change has taken place within the analysed section of data. This would be very important to a more time-sensitive application (on-line analysis) or where the real time detection is not important (off-line detection) but the exact moment of change is needed for further analysis.

2.2. Change Point Detection Methods in Time Series Data

There are well developed methodologies for finding change points in traditional time series data, where a metric is monitored over a number of time bins and evaluated for change. There is a number of methods utilising different data mining techniques which broadly search for abrupt change in the mean or variance of the monitored variables/data. One of such methods, which is used in our experiments, is a Bayesian Change Point (BCP) detection that works under the assumption that the underlying sequence of time series data can be partitioned into a sequence of blocks. Within each of these blocks the data exhibits behaviour described by a set of parameters whose values do not change between blocks. BCP techniques often cite the use of product partition models which are defined based on the assumption that observations within each random partition have independent prior distributions [10]. The number of blocks in the data is unknown and is randomly sampled using the Monte-Carlo technique [9]. The main metric to determine the change event is the posterior probability of change that is equated to an increasing change in a given parameter between the defined bins [11]. [12] is a popular, more recent study that tackles the change point problem from an on-line perspective with time series datasets. The work is based on the previously mentioned assumption that the sequence can be divided into partitions where the places between the partitions are considered as potential change points. The on-line algorithm is constantly updating when new data point is available and after this event the posterior probability of a change is calculated. If this is not considered to be a point of change the computation gets added to a 'run length' which is the time since the last observed change. The probability of a change increases as the run length increases. The calculation of the probability only considers data within a run length.

2.3. Generative Network Models and their applications in Change Point Detection Research

Generative models are usually found in exploratory network analysis and modelling where the goal is to identify interesting structural patterns. They are defined in [16] as a structured probability distribution over entire graphs. In the case of networks, generative models can be used to either produce graph simulations, or ways to represent data in the form of community structures. The benefits come in large networks as they provide an ability to capture and group individual nodes without any prior knowledge of group labels.

Generative network models are the most often used in community detection [17, 18]. We will briefly discuss some research from these areas as a way to describe the different models, but our primary focus is how they are used in change point and anomaly detection problems. We note that there is a clear difference in using models to detect communities than to detect change. In the first case, groups are identified specifically from members of the graph exhibiting common connective behaviour while a change is identified when groups or nodes behaviours no longer conform to the group structures. Of course, a number of approaches have been proposed to detect dynamics of the communities and their evolution over time [19], [20], [21], and [22]. Many of the community detection methods find groups by optimizing selected metric (e.g. modularity [23] or modularity density [24]) and apply this method to different snapshots of a network. Finding differences between those metrics from one time window to another can be use to detect changes in a complex network.

2.3.1. Stochastic Block Models (SBM)

SBMs are one of the most popular generative network models [25]. Wang et al. [15] provides an example of how an SBM can be used in a change point detection. They use the model to infer the communities of the Enron email network. They do not use the model directly to detect change, but as a basis to conduct scan statistics. Scan statistics is a commonly used method in anomaly detection research, which uses the process of 'scanning' smaller sections of the graph to measure the changes compared to recent witnessed behaviour [26].

The problem not addressed in this research, is that although they use dynamic datasets, they use a restrictive application of the SBM representation. The original B group memberships assigned to nodes at the beginning must remain fixed throughout the algorithm, and only the probabilities of the membership matrix are allowed to change over time. We foresee a problem with the inflexibility to account for major structural change as there may be concept drift in the community structure, where an individual vertex or group may change entirely or an entire block may become obsolete if nodes disappear over time.

Karrer and Newman [25] raise another issue with the inference of SBMs, finding in many cases the model lacks the ability to encapsulate the important unique features of different graph datasets leading to radically incorrect structural interpretations. They propose an extension to the model that accounts for the degree distribution in the inference method and find in many cases this is a sufficient improvement on commonly used networks. This extension is supported by other researchers [27] who find in most cases the degree adjusted model should be preferred to the basic version.

A recent study [3] has been conducted using change point detection methods with the degree adjusted block model. These models and associated methods have been praised due to their ability to produce maximum likelihood estimators of the different parameters that characterise the model, which can be used to monitor a change. The ideas for monitoring change are based around considerations of what aspects of change in these parameters would need to be introduced for a graph to exhibit a difference in behaviour. The method is based on three known parameters of an SBM used in [28] for the block model selection problem. Each of the parameters has the ability to model different types of change. For instance the degree parameter (θ) reflects a node's tendency to connect. Changes in the degree parameters have the ability to model changes in the interaction rates of the communities. It is later used in our study as one of the considered global statistics that can measure the overall interaction of the graph nodes. Changes to the community labels have been discussed, though they have not been used in the experiments with real-world datasets, only choosing data where the community labels are known a priori. This is often not the case in real-world networks [25].

There have been many more recent developments that have adapted the SBM to incorporate or to account for different witnessed graph structures and increasing complexity. We will discuss some of these models in detail and their usages in dynamic network research in the next section. The degree adjustment used in the stochastic block model can easily be, and is often, also applied to these other block model extensions.

2.3.2. Mixed Membership and Other SBM's

The major trend in the improvement of anomaly detection in networks is to account for the dynamic nature of community behaviours. Rossi et al. [4] find that accounting for change is key in understanding the dynamics of graph structures. They take inspiration for their graph based dynamic anomaly detection approach from [5], who developed a dynamic and mixed membership SBM model. Mixed membership models allow for a node to become a member of up to all B groups to a different degree represented by a mixed membership vector.

In [4], it has been established that this mixed membership vector is the most important representation of the changes to the graph over time and a model is proposed which tracks this behaviour. This study differs from other CPD methods we reviewed as it focuses more on identifying the time points for individual node anomalies rather than any drastic changes in a whole community structure. Methods taken from [29] are used to establish node roles from the structural features that go beyond the use of membership probability. In many respects the ideas are not wholly suitable for our intended experiments as to identify major events of change we must in principle look at the entire structure. However in spite of such differences it does offer interesting extensions on top of the previous algorithms, such as the model's ability to estimate future behaviour, which performs well in tracking the underlying trend of the data.

Another example of SBM accounting for dynamic behaviour is found in [6] where the layered SBM models are used for modelling complex networks. It is shown that networks do not only possess a single type of pairwise interaction, rather a complete complex system encompasses several layers of interactions that can also help with interpreting changes in time. This paper formulates a generative network model of layered networks that can be generalised for several variants of the SBM incorporating hierarchies, overlapping groups and degree-correction in addition to a layered structure.

Peixoto [6] suggests that dynamic networks should be viewed as a special case where layers start their existence in networks at a specific time based on the value of the edge co–variants. Nodes are assumed to belong to all layers but group membership can depend only on the activity of the group at any given time. They give the opportunity to increase the complexity of the layered model, as degree corrections or mixed membership vectors can be separately specified. They find that the best way to model the structure is in a sequence of small time bins (each bin being a layer) similar to the ideas discussed in section 2.2. The interpretations between layers should be grouped together where there is similar behaviour, and any large differences between these groupings identify a change point. They measure the change in activity by the probability density of an edge being present, which reveals the increases or decreases in activity. The result is a time series sequence of metrics that can then be analysed.

2.3.3. Hierarchical Graph Models

Hierarchical graph models were introduced in [30] and developed further in [31], where the models were used for link prediction. In these studies the hierarchical structure of a graph takes the appearance of a dendrogram, which tries to explain the witnessed behaviour of communities in networks. Higher levels are groups or communities that then split into sub-groups until we reach the lowest levels of individual nodes.

Clauset et al. [31] use the hierarchical structure in the link prediction problem which produces successful results in some cases compared to other popular methods such as a degree product or shortest path [32]. Their conclusion for its success as a predictor is based on the models flexibility to fit to a wide range of network structure types. They find that often group structure models do well at portraying assortative relationships, where groups contain dense connections with few edges between them but struggle when communities have more complex relationships. The inclusion of hierarchy can easily portray more 'dissortative' structures and combinations of both through the relationship portrayal in parent nodes as you move further up the tree.

A unique paper in the field of CPD in complex networks is the study by Peel and Clauset [1], who establish a method close to traditional on-line CPD in time series research. They use a Generalised version of the Hierarchical Random Graph Model (GHRGM) which creates a similar probabilistic dendrogram of the original network structure established in [30]. The main difference between the model employed here is the relaxing of the requirement that the model must produce a full binary tree. Where in the previous structure all nodes must have only two children, in the generalised model a parent can have any number of connections and therefore has the ability to show a model more likened to a block group or community structure.

The method for CPD uses a hypothesis test comparing the interpreted model at a time point within a given window to a null hypothesis model which uses previously witnessed data. They use a sum of edge connection probability for all nodes in the current dataset, finding change points occur when the shape of the estimated probability distribution over a network changes significantly. Peel [1] finds that the Bayesian approach enables the model to learn behaviour adapting to subtle changes as the network evolves. The model is successful when applied to real world networks, where they identify many change points in the Enron graph that can be linked to noted events in the time period. Comparing to the study [15], where block models were used with a combination of scan statistics, we can see that the adaptation to changes in the structure are a clear benefit as they identify many more change points in a time period that is linked well known external events.

Hierarchical models can be considered distinct models in their own right however they are often combined with the SBM to create a more complex expressive structure but have seen no evidence of the use of this combined version of the model for change point detection or anomaly detection problems.

2.3.4. Critical Review of Generative Network Models

We have discussed benefits of the generative models in CPD and other research topics through their ability to portray important structural features relevant to the problem space. Here, we address challenges that are often highlighted in the literature related to choosing and inferring generative model structures.

Jacobs and Clauset [16] claim that the model selection for generative-network models remain an open challenge not only in the context of our change point topic, but in the usage of models as a whole. For instance the decision of when to use an SBM with or without degree adjustment is still debated. They find many proposed methods to decide between models do not always work, and that failure may be due to inappropriate assumptions made.

Peixoto [27] addresses the problem that more complex models are more favoured by some statistical performance tests due to the over-fitting. They propose a method that aims to choose the best model fit while minimising the amount of parameters needed to be estimated. This method produces an entropy figure that can be used to measure description length of the model fit to the data. A lower entropy figure would signify that the model is better fit for the data. It is also found using this entropy measure that often more complex block models (e.g. with overlapping MM groups) are better only in a small number of cases. On the other hand, the previously mentioned block models with a degree correction are almost always favoured. They conclude it might be the case that individual node properties are equally important as group/block connection properties in the case of network formation.

In [16] authors relate the problem of model selection to that witnessed in general clustering algorithms. [33] suggests that the choice of clustering algorithms need not be chosen solely on the results of a significance test but is also related to the end use of the clustering application. [1] discuss the ability for their algorithm using the HRG model to be replaced by any other probabilistic model structure. They also offer the view that the HRG is very suitable in the CPD problem domain due to the previously mentioned benefits of adaptability to all kinds of complex network structures. By comparing the performance of two different types of models we might be able to discover if there are properties in the networks or in the problem space that make certain types of models better than others.

In some cases in change point or anomaly research there is a failure to address the reasoning behind the model selection compared to the other possible interpretations available. This may possibly be due to the still early stages of such research and authors are searching for general solutions rather than problem specific applications. This presents an opportunity to measure the suitability of generative models when applied to different real world problems. By conducting controlled change point detection experiments on different types of real world network datasets modelled with different generative models we may find conclusions on the matter if choice of model effects the ability to identify changes in dynamic graphs.

2.4. Summary of the literature

When analysing changes, an obvious first step would be a review of metrics that give an overall understanding of the graph behaviours like some of those used in our experimental analysis. Previous research suggests [34] that metrics, such as density and clustering coefficient, tend to be stable over time. It would be interesting to view how these metrics perform in a change point setting compared to the complex topology metrics that can be extracted with generative models. It might be the case that overall properties offer too broad a view, and are not able to account for the more subtle changes in group dynamics that generative models have been praised for.

The previous research reviewed in section 2.3.4 has highlighted some concerns when working with generative network models, which need to be overcome to effectively use them. One of the key issues is their computational complexity which combined with the requirements of dynamic analysis of large, complex networks may render them infeasible to apply. It should be noted that the global network statistics discussed, among others, in [34] are much more computationally efficient. Additionally there is no preferred method for choosing between generative models in a change point detection problem which we have attempted to address in our study.

The main reason behind the choice of generative models in change point detection is their effective way of representing communities in a probabilistic manner. The parameters produced by SBM's can be effectively monitored for changes. These for instance include edge connectivity matrices and overall number of blocks. However, we found that the datasets that are often used have community labels already known or alternatively remain fixed during the analysis. Leaving labels fixed over dynamic graphs may be too restrictive to learn about change points if major concept drifts happen in the analysed period. The complexity of SBMs particularly for the more developed extensions reviewed in sections 2.3.2 and 2.3.3, have a tendency to over-fit, and this should be analysed.

The method established in [1] is one of the more developed CPD techniques for a generative model providing a clear methodology based on traditional time series Bayesian change point analysis. It also avoids the problem often found in block model research and does not force any fixed labels on to nodes. The analysed metric for change, the sum of all edge probabilities will therefore be used in our analysis.

Many of the traditional change point detection techniques were utilised and adopted for the change analysis in networks. Due to the difference in data structure there is a potential loss of important information by transferring the network structure (rich representation) into a single metric (simplistic representation of a complex phenomena). However, as in this study we are interested in extending our knowledge into which of many possible graph metrics are the most suitable for modelling changes in networks and particularly the change point detection task, the identified approaches used in time series analysis have been found to be appropriate to our goals.

3. Selected metrics for change point detection

This section presents the selected network statistical indicators and generative models' parameters that have been used in the experimental analysis. The justification for selection together with a brief descriptions are provided below. Summary of those metrics, together with formulas used to calculate them, is presented in Table 1.

3.1. Network Properties to Analyse the Network Structure

The used global metrics related to the network structure are:

– **Network Size (N)**: is the number of active nodes during the given time period. This was used to give an indication if there is any shift in overall network size and point to any potential major shifts in groups (or blocks) that could be uncovered in generative models. For instance in the case of many disappearing nodes, this could have a direct effect on group dynamics.

– **Average Clustering Coefficient (ACC)**: The clustering coefficient gives an indication of how closely connected a graph is. It could be assumed when applying this to change point detection that a shift in the clustering coefficient would indicate a change in the dynamics of the graph. The global method explained in [35] is used to calculate this metric.

– **Density (D)**: The network density is a metric that represents the ratio of the number of edges in the graph to the total number of potential connections. So, similarly to the network size, changes in number of edges may indicate evolution of communities.

– **Average Shortest Path (ASP)**: The average shortest path measures the smallest number of connections between all potential pairs of nodes in the graph. This again gives an idea of the connectivity of the graph and the closeness of groups ([35]).

3.2. Generative Network Models & Parameters for the Analysis of Network Structures

Although there have been a number of successful approaches in the area of change point detection in complex networks that model the problem using variations of Stochastic Block or Hierarchical Block Models, no comparative study was performed to determine why one should be preferred over another. For this reason this research utilised two variations of block models to allow for comparison of the techniques. The selected models are:

* **The degree adjusted Stochastic Block Model (SBM)**. For the SBM inference the primary task after group creation is to create the membership matrix, which is the inferred sum of edge-counts between and within groups. Please note that in the literature a representation of the block matrix that is more commonly used is the probability that two nodes in different groups will connect [6]. However, the implementation used in our experiments employed edge-counts as opposed to probability, and has been found to be equally rep-

resentative to the probabilistic method. The membership matrix is estimated once group assignments have been equated and is randomly sampled and produced with the MCMC algorithm. For a degree adjusted model this includes the additional parameter K on to each vertex, which is the degree sequence of a node. This provides a restriction on group memberships by applying an average degree target that each group must obey. This has the effect of ensuring high-degree nodes are more likely connected to low degree nodes in a group.

* **The variation on the Generalised Hierarchical Random Graph (HRG) model**. In the model used for the experiments each layer in the nested structure is a distinct block model. The nodes in the higher levels represent the multiple groups in the lower levels, and their connections. As in Peels model [1], this is more flexible to a hierarchical random graph, as all nodes are present in the structure and no restrictions are placed on the number of children a parent node can have. Please note that the model is again, like with the block model, not probabilistically inferred like in Peels method, but the connectivity is determined by an edge count.

Parameters extracted from the above two types of generative models to be analysed in the CPD setting and a brief discussion of why they were included are described below.

– **Entropy ($\mathcal{S}$)**: is used to measure the description length of the model, given the current data structure. It has been developed as a way to choose between different model structures [36]. Entropy's value is extracted for each time point and could reveal a change in the complexity of the model as the graph structure changes. It is derived from the size of network and any increases in the count of edges or nodes will affect the entropy figure. Other effecting values are the number of blocks and connectivity matrix, the more complex the decided graph structure, e.g. a larger number of blocks with 'dissortative' connection probabilities (edge counts) would assume a larger description length and therefore a larger entropy value.

– **No. of Blocks (B)**: The number of blocks (B) in a network dataset refers to the number of communities established by the inference methods using the

Table 1: Measures for change point detection used in the experiments

Metric	**Formula**
Network Properties	
Network Size	N
Average Clustering Coefficient	$ACC = \frac{1}{N}\sum_{i=1}^{N}\frac{2M_i}{k_i(k_i-1)}$
Density	$D = \frac{2E}{N(N-1)}$
Average Shortest Path	$ASP = \sum_{s,t\in N}\frac{d(s,t)}{N(N-1)}$
Notation: N – no. of nodes; M_i – no. of connections between neighbours of node i; k_i – degree of node i; E – number of edges	
Generative Network Models Parameters	
Entropy	$\mathcal{S} = -\ln P(G\vert\theta) - \ln P(\theta)$
Number of Blocks	B
Edge Probabilities	$\gamma = \sum lnP(G_t\vert T_t, \theta)$
Hierarchy Layer Membership	$HLM = \sum L \times B$
Notation: G – graph; θ – the model parameters of membership matrix made up of B number of blocks and E number of edges between them; G_t – graph at time point t; T_t – model used at time t (block or hierarchical structure); L - layer figure	

recorded connectivity patterns [37]. The number of blocks is randomly sampled and calculated by determining the best fit using the additional model parameters (the connectivity matrix). The number of groups is likely to change over time and major changes from the previously observed behaviour. For example a huge increase in block numbers could signal an increasingly random network or a more dense network structure.

– **Edge Probabilities (γ)**: Edge probability is the likelihood that two individual actors in the network will connect. For a block model this is based on the connectivity matrix and in the case of hierarchical models this is determined through the hierarchy levels. Hierarchical models have most often been used in edge prediction methods [31]. This probability is utilised in [1] to determine change points in their on-line algorithm by calculating the sum of the probabilities of all the current pairs in a time window, normalised by the sum of edge probabilities given the historical data.

– **Hierarchy Layer Memberships (HLM)**: The HRG model has a parameter, which is calculated based on a number of layers in the hierarchy and number of groups within each layer. To determine how many groups are at each level, a separate block model is run on each layer. This structure could also be monitored over the time period in a similar way to the number of blocks in an SBM. HLM allows us to capture any changes in the blocks and layer structural complexity.

4. Experimental Analysis

4.1. Datasets & Data Preparation

The experiment was conducted on two different datasets that are often used in network change point and anomaly detection research. Networks extracted from those datasets are: (i) MIT reality mining social proximity network [38] and (ii) Enron email network [39]. In order to analyse the dynamics of selected graph metrics and parameters of generative models both networks were split into time windows. When choosing size of time window there is a risk that the change may not be fully developed if the time window is too small. On the other hand, a change can be completely overlooked if too large a window is selected. To be comparable with other studies (e.g. [1]) the datasets were split up into weekly time windows.

4.1.1. MIT Reality Mining Network

MIT reality mining social proximity network was first collated and described in [38]. This dataset contains data from 100 students and staff using blue tooth to study the social patterns of interaction over the course of nine months. This dataset has a number of known groups, such as freshmen or staff, however we choose to use the modelling search technique to establish groups outside of these pre-defined labels. The known events in this dataset correspond to common events in most school terms, such as exam periods, Christmas and spring breaks. Figure 1 shows the change points in MIT network (nine in total) against which the selected metrics are tested during the experiments. As an example, in Figure 1, the line represents the changes in ASP over time.

4.1.2. Enron Email Network

The second dataset used was the Enron Email Network [39]. The network is created from a collection of emails in the Enron company from 1999–2002 which contains over 600,000 emails from the inbox of 151 different users. Relationship edges represent an email being sent to or from different accounts. This dataset is often used to detect anomalies due to a number of events that happened that eventually led to the company filing for bankruptcy. Due to computational constrains, we used a smaller portion of the dataset, including years 2000–2002. We have estabished a set of events that includes e.g. major company announcements, stock price increase, mass redundancies and CEO changes that were available at [40]. A chart showing these events (31 in total) plotted against the number of active nodes in the graph during the corresponding week is shown in Figure 2.

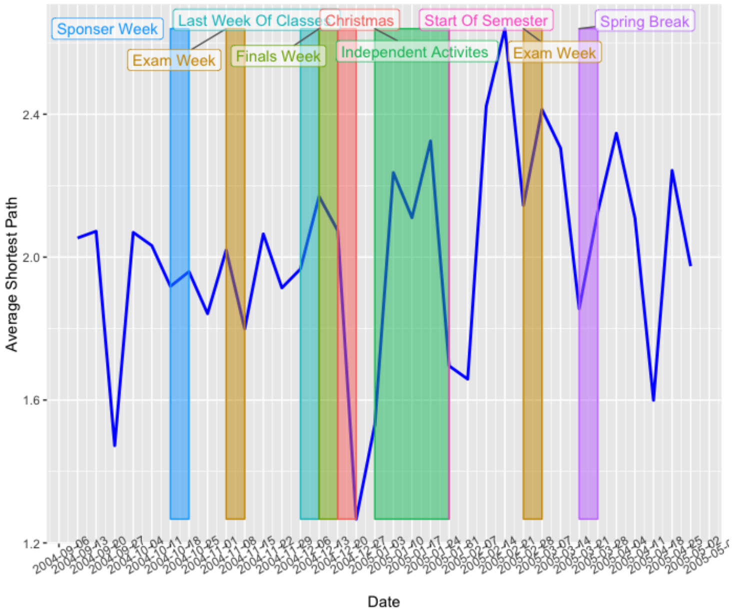

Figure 1: ASP over time in MIT network; also showing change point labels

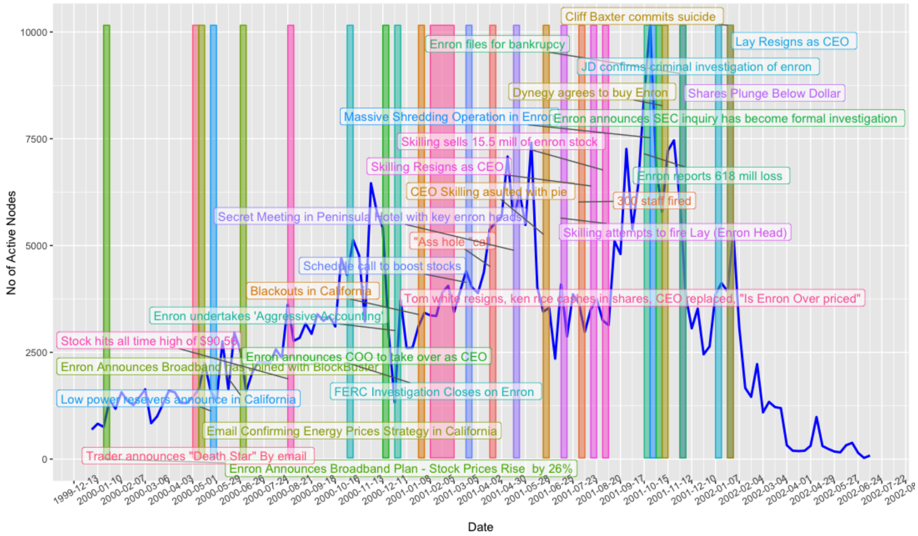

Figure 2: Active Nodes over Time Enron Dataset Jan 2000–June 2002: Including Change Point Labels

4.2. Experimental Set Up

Both datasets were cleaned and divided into time windows of size one week. The following steps were repeated for each of the time windows:

- Create graph representation of the data for windows t: G_t.
- Extract Global Metrics from G_t: number of active nodes (size of the network), average clustering coefficient (ACC), density, average shortest path (ASP).
- Create Generative Model $M_t = P(G|\theta)$ of Graph G_t using MCMC sampling algorithm (for both SBM and HRG).
- Extract the Generative Metrics from the M_t of each model. This includes: Entropy, Sum of Edge Probability, No. of Blocks, and Hierarchy Layer Membership (Layers).

Results obtained using this procedure were analysed. First, the enumerated metrics were discussed and correlation between them investigated. The second part of analysis aimed at evaluating the ability of those metrics to measure change. After the metrics were collected at pre-established time points using the above process, the results produced a dataset of all given metrics. The next step was to establish a way to determine if these metrics provide a good indication of changes during these time points. The validation part of the method needed to compare the extracted results against the ground truth known events. In order to establish which changes were detected by the network metrics, first a Bayesian change point analysis was performed on each graph metric using Bayesian off-line change point detection [41]. This technique was chosen as it provides an output values between 0 and 1 rather than the restrictive change or no change labels that other change point methods use. By using the probability output, it allowed us to use our own threshold. A time period was classified as a change point if the posterior probability value was greater than the average posterior probability in the dataset for a given metric. This was then compared to a binary sequence representation of known events (where 1 indicates an event, 0 no event).

The performance of all the explored metrics; Size, Density, ACC, ASP, Entropy, Sum of Edge Probability, Number of Blocks and Hierarchy Measure against the known external change points is calculated using the Precision and Recall approach [42]. A point was classified as a true positive if the posterior probability is above specified threshold and a ground truth event has occurred within a varying time period of +/- 0 to +/- 2 weeks. Those results show the top performing metrics and allow to establish which metrics are related to change. Additionally, the computational time of the experiment is analysed. It is an important consideration as it provides a view on whether the notoriously time heavy generative models offer more value to the results obtained by using traditional global statistics which are comparatively much easier to calculate.

4.3. Descriptive Statistics and Correlation Matrices

Metrics calculated for each time window were firstly analysed with the use of descriptive statistics. The analysis is supported by a correlation matrix between all the metrics. We then move onto a review of the performance of each metric in the context of a change point detection setting. This was measured using the Precision and Recall rates calculated from the posterior probability of a change in the distribution of the network metrics, against a set of classified known events. We also consider the computational cost incurred when using generative models.

4.3.1. Results

As described in the Experimental Setup section, each of the chosen metrics (Size, Density, ACC, ASP, Entropy, Sum of Edge Probability, Number of Blocks and Hierarchy Measure) was extracted from a network snapshot at weekly intervals for the two datasets (MIT and Enron). This created a collection of time series data which were analysed using descriptive statistics (Tables 2 and 4).

Table 2: Descriptive Statistics: MIT Network

Statistic	Mean	St. Dev.	Min	Max
No. Nodes	61.56	18.72	8	85
No. Edges	416.09	271.40	8	941
Density	0.20	0.06	0.12	0.30
ACC	0.51	0.08	0.33	0.60
ASP	2.01	0.30	1.27	2.64
SBM: Blocks	32.15	15.78	3	59
SBM: Entropy	8,016.33	5,668.24	185.95	18,294.35
SBM: $\sum$ Probs.	−482.65	308.19	−1,155.71	−14.34
HRG: Levels	506.09	308.64	50	1,017
HRG: Entropy	4,754.34	2,887.94	158.51	10,320.35
HRG: $\sum$ Probs.	−466.26	294.62	−1,059.19	−9.17

Table 3: Descriptive Statistics: Correlation Matrix of MIT Network Metrics

	HRG:$\sum$ Prob	HRG:Entropy	HRG:Layers	No. Nodes	No. Edges	Density	ACC	ASP	Events
HRG:$\sum$Prob	1	-0.984	-0.960	-0.889	-0.993	-0.484	-0.666	0.107	-0.141
HRG:Entropy	-0.984	1	0.970	0.871	0.991	0.498	0.664	-0.100	0.156
HRG:Layers	-0.960	0.970	1	0.864	0.967	0.468	0.630	-0.078	0.086
No. Nodes	-0.889	0.871	0.864	1	0.874	0.079	0.552	0.151	0.137
No. Edges	-0.993	0.991	0.967	0.874	1	0.508	0.682	-0.106	0.157
Density	-0.484	0.498	0.468	0.079	0.508	1	0.501	-0.475	0.215
ACC	-0.666	0.664	0.630	0.552	0.682	0.501	1	-0.022	0.231
ASP	0.107	-0.100	-0.078	0.151	-0.106	-0.475	-0.022	1	-0.193
Events	-0.141	0.156	0.086	0.137	0.157	0.215	0.231	-0.193	1

Table 2, shows a summary of the metrics for the MIT dataset, and Table 3 shows a correlation matrix of these metrics. The descriptive statistics show that MIT is a highly connected graph. The average number of blocks extracted over the time period was relatively high compared to the number of nodes in the network, suggesting that the graph is made up of mostly small communities. The ASP is small and the ACC measure on average is 0.506 which suggests a highly connected small-world like network.

The Correlation matrix for MIT metrics reveals two different groups of metrics. Firstly generative models and network size indicators (no. of nodes and edges) which are all very well correlated. The second group, the remaining global metrics, have little to no linear relationship with the first group or each other. Looking at the descriptive statistics (Table 2) one can see that comparably the global graph metrics have a lower standard deviation (std) then the others which suggest a more stable trend over time.

Table 4: Descriptive Statistics: Enron Data

Statistic	Mean	St. Dev.	Min	Max
No. Nodes	3,068.773	2,031.773	27	10,157
No. Edges	5,228.833	3,963.222	31	20,148
Density	0.003	0.009	0.0004	0.088
ACC	0.131	0.039	0.042	0.246
ASP	1.361	0.131	1.136	1.760
SBM:Blocks	18.826	11.340	1	56
SBM: Entropy	46,405.460	35,329.750	367.955	187,670.200
SBM: $\sum$ Probs.	23,755.490	19,122.730	190.397	94,406.040
HRG:Layers	318.598	292.994	5	1,466
HRG: Entropy	45,206.980	34,046.610	367.955	179,330.500
HRG: $\sum$ Probs.	23,144.730	18,515.700	190.397	91,508.410

Table 4, shows a summary of the metrics for the Enron dataset, and Table 5 shows a correlation matrix of these metrics. The descriptive statistics show that the Enron network is a sparse structure, reflected in the average number

Table 5: Enron: Correlation Matrix

	SBM: $\sum$ Prob	SBM: Entropy	SBM: Blocks	No. Nodes	No. Edges	Density	ACC	ASP
SBM: $\sum$ Prob	1	-0.978	-0.884	-0.974	-0.983	0.310	-0.048	0.639
SBM: Entropy	-0.978	1	0.928	0.988	0.994	-0.323	0.043	-0.658
SBM: Blocks	-0.884	0.928	1	0.922	0.921	-0.358	0.079	-0.664
No. Nodes	-0.974	0.988	0.922	1	0.987	-0.365	0.013	-0.673
No. Edges	-0.983	0.994	0.921	0.987	1	-0.325	0.091	-0.658
Density	0.310	-0.323	-0.358	-0.365	-0.325	1	-0.136	0.167
ACC	-0.048	0.043	0.079	0.013	0.091	-0.136	1	-0.006
ASP	0.639	-0.658	-0.664	-0.673	-0.658	0.167	-0.006	1

of blocks for this dataset relative to the number of nodes. The average number of active nodes is over 3,000 but the number of communities is only 19, which means that the graph contains a small number of large groups. The peak in group size is also aligned with peaks in the number of nodes, edges and all other generative models metrics. This peak is around the time that Enron begins shredding information and eventually files for bankruptcy. The two entropy measures for the HRG and the SBM have very similar values (see Table 4. The HRG entropy figure is slightly lower and would suggest that this model is a better fit for the dataset.

The Correlation matrix shows that the generative models and metric size indicators are all correlated, and the global metrics have little or no correlation with any other metrics with the exception of the ASP that has a higher linear correlation (above 0.6) with the generative models metrics. The log descriptive statistics (Table 4) shows that this second group of metrics has a much lower std from the mean than the generative and size metrics, which suggest a more stable trend over time.

4.3.2. Discussion

Despite the clear differences in the networks structures, in terms of size, community structure and density there are many common characteristics. The complexity measures for the generative models, Entropy and the No. Blocks (or layers for the Hierarchical Model), revealed that the generative model structure experienced many changes over time. The block labels are often kept fixed during dynamic network analysis. However, the results support the view that fixing group labels may put unwanted restrictions on the models. The Entropy metric was originally developed in [6] to help decide between model adaptations, so we can also conclude from Tables 2 and 4 that according to this theory the Hierarchical Model is the preferred model for both datasets. This agrees with the literature that the HRG has the ability to provide a better fit for more complex dissortative relationships that are most often the case in real world graphs.

It is also apparent from Tables 2 and 4 that the HRG and SBM have very similar average and range of values for both Entropy and Probability Sum Indicators. The correlation tables find that the HRG and SBM edge probabilities are highly correlated with each other (see Table 6 for correlation between generative metrics). The SBM and HRG entropy have a correlation of 1, suggesting that the sampling algorithm produces representative samples. The three generative metrics (for both the SBM and the HRG) are highly correlated with the number of edges in the graphs for both datasets. The reason for this is probably that all of the metrics heavily rely on the number of edges in their parameter estimation.

In [43] authors show that global graph metrics often remain stable over time. Our results agree strongly with this conclusion, as looking at Table 4 two of our global metrics (ASP and ACC) have the lowest std figures by far in comparison to the other metrics. The exception to this is the Density metric, however this figure is stable throughout the time and was only subject to variance at the beginning and end of the experimental time period. The correlation Tables

Table 6: Enron: Generative Model Metrics Correlation Matrix

	SBM: No of Blocks	SBM: Entropy	SBM: $\sum$ of Edge Prob.	HRG: Layer complexity	HRG: Entropy	HRG: $\sum$ of Edge Prob.
SBM: No of Blocks	1	−0.983	−0.872	0.987	−0.982	−0.896
SBM: Entropy	−0.983	1	0.896	−0.979	1.000	0.927
SBM: $\sum$ of Edge Prob.	−0.872	0.896	1	−0.870	0.900	0.905
HRG: Layer complexity	0.987	−0.979	−0.870	1	−0.978	−0.884
HRG: Entropy	−0.982	1.000	0.900	−0.978	1	0.928
HRG: $\sum$ of Edge Prob.	−0.896	0.927	0.905	−0.884	0.928	1

Table 7: MIT: Total Identified Change Points

Metric	No. Change Points	Threshold
Nodes	7	0.159
Edges	8	0.106
Density	7	0.147
ACC	7	0.081
ASP	7	0.083
SBM: No. Blocks	9	0.236
SBM:Entropy	6	0.169
SBM: $\sum$ Probability	9	0.102
HRG: Layers	7	0.153
HRG: Entropy	10	0.129
HRG:$\sum$ Probability	9	0.094

3 and 5 also reveal that our global graph metrics (ASP, ACC and Density) have a smaller degree of relationship to the other metrics, as the correlation between them tend to be below 0.5. In the MIT dataset the ACC has a stronger relationship (above 0.6) with the generative model metrics and number of edges. In the Enron correlation matrix the ASP shows a stronger relationship with the size and generative metrics of (again) above 0.6.

This section gave us a general understanding of the trends over time for each calculated metric suggesting that they may be useful to measure change. For instance many of the metrics for the MIT dataset drop (or peak in the case of the Sum of Edge probabilities) during the Christmas period. In the Enron graph a similar peak happens during the Mid Sep-Mid Oct 2001, in the weeks prior to Enron filing for bankruptcy.

4.4. Analysed Metrics as Indicators of Change

4.4.1. Results

The validation results for each metric are shown in Tables 8 and 10. Firstly focusing on the MIT results in Table 8, the best performing metric was the HRG:$\sum$P with a Precision score of 0.55, and one of the highest Recall of 0.625. It is the best in terms of accuracy when the window size is 0. Once the window

Table 8: MIT:Precision (P) & Recall (R) Results ((n) = window size)

Metrics	P	R	P(1)	R(1)	P(2)	R(2)
Nodes	0.29	0.25	0.86	0.27	1	0.26
Edges	0.37	0.37	1	0.36	1	0.30
Density	0.29	0.25	0.71	0.23	0.86	0.22
ACC	0.29	0.25	0.71	0.23	0.71	0.18
ASP	0.14	0.12	0.86	0.27	0.86	0.22
SBM: Blocks	0.22	0.25	0.78	0.32	0.78	0.26
SBM:Entropy	0.33	0.25	0.83	0.23	0.83	0.18
SBM: $\sum$ Prob	0.33	0.37	1	0.41	1	0.33
HRG: Layers	0.29	0.25	0.86	0.27	0.86	0.22
HRG: Entropy	0.30	0.37	0.80	0.36	0.90	0.33
HRG: $\sum$ Prob	0.56	0.62	1	0.41	1	0.33

size is increased to +/- 1 the Precision and Recall scores improve. Both measures of the probability sum are able to capture all the change point values. A notable top performing metric is also the number of edges that is able to capture all the change points correctly within a window size of +/-1 week. The other generative model metrics, ACC, ASP and density all have similar results in terms of Precision. For the largest window size, many of these metrics are not able to improve their Precision scores from the window size of +/-1 week. The Recall figure is low for the remaining metrics, as expected due to the impact of the window size. However, it can improve even with a large window, suggesting that the gains in true positives rates outweigh the increase in false negatives.

From the mapping of change point probability over time (Figure 3) for each metric, one can see that the biggest change point discovered was around the week of Christmas. The other change points correctly identified are difficult to see in Figure 3, as they were of a much lower probability but still above the threshold (which was set at above the average probability for the metric displayed in Table 7). The total number of detected change points by different metrics is shown in Table 7. The HRG entropy has identified the most change points, more than the total number of events (which was 9). The SBM entropy metric identified only 6 change points, the lowest of all results and this also

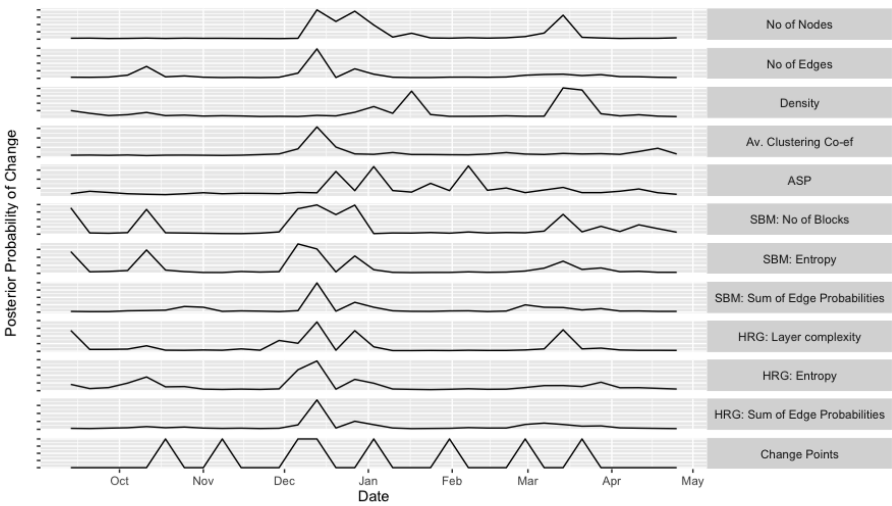

Figure 3: MIT: Posterior Probability of A Change Over Time for all Metrics. All y axes have ranges between 0 and 1.

resulted in a low recall figure.

The Enron's Precision and Recall results (Table 10) are more balanced than for the MIT dataset. None of the metrics performed much better than the others. There is a clear distinction between the performance of Global metrics (excluding the size indicators) and the Generative Models. The best metric for precision was the number of Nodes, however only 35% of the total change points have been identified. When the window size was increased to +/-1 there was some improvement (to over 60%) for the Number of Nodes, Edges, SBM:Entropy and HRG:Probability. For this dataset density was the worst performing metric, followed by the ACC and ASP, even when the window size is increased to the highest level (+/-2 weeks). The Recall rates are all equally bad for any given metric. The Enron dataset had a lot more change points to be identified, which are quite evenly spread out. The change points are also not cyclical known behaviours but one off events that are difficult to tie down to exact dates and times. Considering this the metrics perform very well and identify many of the events and changes in a relatively small window of +/-1 week. When reviewing the change points against the ground truth events in Figure 4, one can see there

Table 9: Enron: Total Identified Change Points

Metric	No. Change Points	Threshold
Nodes	29	0.148
Edges	32	0.17
Density	20	0.138
ACC	29	0.07
ASP	32	0.09
SBM: No. Blocks	32	0.168
SBM:Entropy	28	0.175
SBM: $\sum$ Probability	27	0.16
HRG: Layers	35	0.232
HRG: Entropy	27	0.171
HRG:$\sum$ Probability	32	0.182

Table 10: Enron: Precision (P) & Recall (R) Results ((n) = window size)

Metric	P(0)	R(0)	P(1)	R(1)	P(2)	R(2)
Nodes	0.34	0.36	0.62	0.27	0.82	0.28
Edges	0.31	0.36	0.62	0.30	0.84	0.31
Density	0.10	0.07	0.25	0.08	0.30	0.07
ACC	0.21	0.21	0.48	0.21	0.62	0.21
ASP	0.22	0.25	0.47	0.23	0.59	0.22
SBM: Blocks	0.22	0.25	0.56	0.27	0.81	0.30
SBM:Entropy	0.29	0.29	0.61	0.26	0.82	0.27
SBM: $\sum$ Prob	0.30	0.29	0.56	0.23	0.85	0.27
HRG: Layers	0.23	0.29	0.60	0.32	0.89	0.36
HRG: Entropy	0.26	0.25	0.59	0.24	0.81	0.26
HRG:$\sum$ Prob	0.31	0.36	0.62	0.30	0.81	0.30

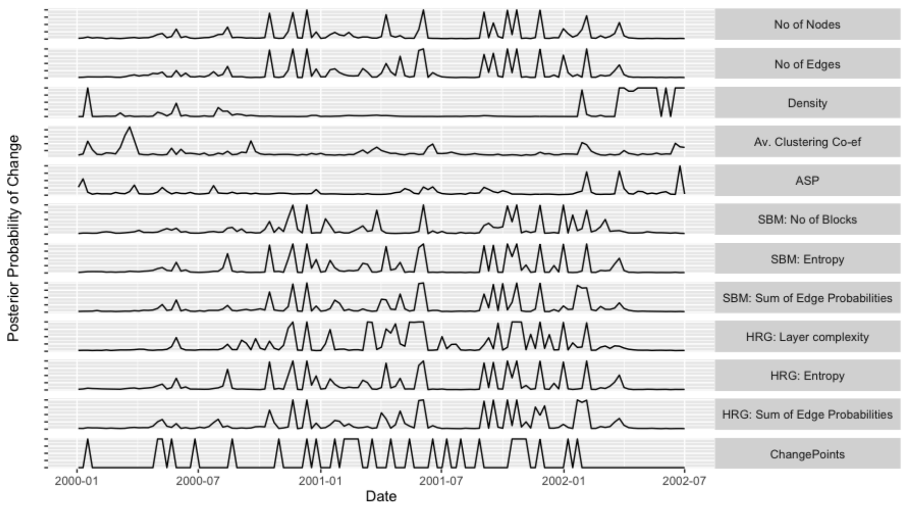

Figure 4: Enron: Posterior Probability of A Change Over Time for all Metrics. All y axes have ranges between 0 and 1.

are many change points of high probability. There are some change points that none of the metrics capture during Feb to Jul 2001.

4.4.2. Discussion

In the case of the Enron dataset the HRG performed better in terms of both Precision and Recall when the window is small. But when the window increases in size, the SBM metrics have higher Precision, therefore indicating they were a better representation of the ground truth. Looking at an overall number of change points for Enron one can see the HRG had a higher number of total change points (on average), which could indicate that this model can better detect small underlying changes. As just because points are not in the pre-defined list, it does not mean conclusively that a change did not occur. For the MIT dataset, the HRG resulted in the best scores. It may be interesting to explore this result further, to discover what features present in this dataset could be an indicator that it is suited to change point analysis. Testing this metric on other similar datasets, either in subject or density structure, could test if this model works well for this type of data. We were unable to conduct this study due to a lack of computational resources. There is also a lack of dynamic datasets of a reasonable size and with known events.

One of the clear conclusions is that the number of edges and nodes gave more than adequate results for both datasets compared to the computationally expensive generative models. The total time taken to produce Hierarchical graph metrics for the Enron datasets was 6.4 days, with run times varying from 30–68 minutes per model whereas counting Number of Nodes and Edges takes seconds. Both Nodes and Edges are heavily correlated with all the generative model metrics and have the ability to portray the structural change just as well in the case of the Enron dataset. When it comes to portraying many changes the generative models were able to discover more change points in total for the Enron data structure, however the difference was minimal. For the MIT dataset all changes were captured with only a small window of error by the no. of nodes and edges. When thinking about the structural changes that would go on over a school term, this would mostly consist of people going home for winter and summer, that would correspond to a change in these two metrics. Major changes to the community structure would probably be more applicable to the Enron graph which had few very large communities. This may be the reason why more changes were identified by the community centric generative model metrics.

5. Conclusions and Future Work

The aim for this research was to determine what available metrics and commonly used descriptive statistics are best at revealing changes in complex networks. The metrics (see Section 3) were extracted from weekly time windows of two network datasets (MIT and Enron). These datasets are both commonly used in change point detection studies and are of different characteristics and sizes.

Comparing the results between two models (Stochastic Block Model and Hierarchical Block Model) reveal that all the generative metrics followed a similar trend over time. Both the SBM and HRG metrics were strongly correlated to the size metrics in the graph, and their entropy figures had been perfectly correlated with each other. The hierarchical model did appear slightly more sensitive to change as the metrics on average identified more change points than the SBM. Also the average entropy figure for the HRG provided a better fit on both datasets used in the experiment. The major downside to this model, however, is that it is computationally heavy.

Despite the successful results from the generative models, the computational time taken to create them became a pointless effort when reviewing the results. The final results showed the simplest metrics, **number of edges and nodes in the network**, in most cases performed just as well at detecting changes as the generative model metrics across both datasets. It can also be concluded that Density, ACC and ASP were all poor detectors of changes, as they have a tendency to remain stable over time. They identified many changes overall, but these were often not related to the ground truth change points.

The generative model metrics are inferred from the number of nodes and edges, so it makes sense that they correlate so strongly and produce similar results. The results conclude that the hierarchical block structure offered only slightly better representations of structural change for the two datasets. When considering the computational effort there is no reason that generative models should be preferred over the simple node or edge count if the primary aim is to discover or monitor global structural change points, especially in larg networks.

This research has pointed to a number of areas that could be further expanded. One area would be changing the time granularity, to assess if a smaller time granularity could change the results, e.g. using daily interaction rates. This would be interesting for the Enron graph, around events of particular notoriety (such as the point at which they file for bankruptcy). The natural extension would be to repeat the experiment on more datasets. As this would help test the reliability of the number of edges and nodes as change point indicators. It would be insightful to see if any data with similar properties give similar results. This then has the potential to produce an on-line algorithm that could detect shifts in these measures. We also briefly discussed the idea of exploring more types of block models, namely the mixed membership SBMs. However, in the light of the findings and the minimal difference in performance of the two model structures tested, it would not be the highest priority.

6. Authors' Bios

[]Lucy Kendrick works at JP Morgan. Bournemouth as data analyst and developer. She obtained her BA in Finance and Economics in 2014 from Bournemouth University. In 2016 she obtained MSc in Applied Data Analytics from Bournemouth University. .

[]Katarzyna Musial received her MSc in Computer Science from Wroclaw University of Technology (WrUT), Poland, and an MSc in Software Engineering from the Blekinge Institute of Technology, Sweden, both in 2006. I was awarded PhD in November 2009 from WrUT, and in the same year I was appointed a Senior Visiting Research Fellow at Bournemouth University (BU), where from 2010 I was a Lecturer in Informatics. I joined Kings in November 2011 as a Lecturer in Computer Science. In September 2015 I returned to Bournemouth University as a Principal Academic in Computing where I was also a member of Data Science Institute. In September 2017 I joined University of Technology Sydney as an Associate Professor in Network Science in School of Software and I am also a member of Advanced Analytics Institute.

[ ]Bogdan Gabrys is a Data Scientist, a Professor of Data Science and a Director of Advanced Analytics Institute at the Faculty of Engineering and IT, University of Technology Sydney, Australia. His research, consulting and advisory activities have concentrated on the areas of data science, complex adaptive systems, computational intelligence, machine learning, predictive analytics and their diverse applications. In particular, he has pursued the development of various statistical, machine learning, nature inspired and hybrid intelligent techniques especially targeting data and information fusion, learning and adaptive methods, multiple classifier and prediction systems, processing and modelling of uncertainty in pattern recognition, diagnostic analysis and decision support systems. He is an accomplished author (with over 150 publications), frequently invited speaker at international events and fora and sought after data science expert. More information can be found at: http://bogdan-gabrys.com/

References

[1] L. Peel, A. Clauset, Detecting change points in the large-scale structure of evolving networks, arXiv:1403.0989.

[2] L. Akoglu, H. Tong, D. Koutra, Graph based anomaly detection and description: a survey, Data Mining and Knowledge Discovery 29 (3) (2015) 626–688.

[3] J. D. Wilson, N. T. Stevens, W. H. Woodall, Modeling and estimating change in temporal networks via a dynamic degree corrected stochastic block model, arXiv:1605.04049.

[4] R. A. Rossi, B. Gallagher, J. Neville, K. Henderson, Modeling dynamic behavior in large evolving graphs, in: Proceedings of the sixth ACM in-

ternational conference on Web search and data mining, ACM, 2013, pp. 667–676.

[5] W. Fu, L. Song, E. P. Xing, Dynamic mixed membership blockmodel for evolving networks, in: Proceedings of the 26th annual international conference on machine learning, ACM, 2009, pp. 329–336.

[6] T. P. Peixoto, Inferring the mesoscale structure of layered, edge-valued, and time-varying networks, Physical Review E 92 (4) (2015) 042807.

[7] W. Wang, P. Jiao, D. He, D. Jin, L. Pan, B. Gabrys, Autonomous overlapping community detection in temporal networks: A dynamic bayesian nonnegative matrix factorization approach, Knowledge-Based Systems 110 (2016) 121 – 134.

[8] M. Basseville, I. V. Nikiforov, et al., Detection of abrupt changes: theory and application, Vol. 104, Prentice Hall Englewood Cliffs, 1993.

[9] D. Barry, J. A. Hartigan, A bayesian analysis for change point problems, Journal of the American Statistical Association 88 (421) (1993) 309–319.

[10] D. Barry, J. Hartigan, Product partition models for change point problems, The Annals of Statistics (1992) 260–279.

[11] X. Xuan, K. Murphy, Modeling changing dependency structure in multivariate time series, in: Proceedings of the 24th international conference on Machine learning, ACM, 2007, pp. 1055–1062.

[12] Y. Kawahara, M. Sugiyama, Change-point detection in time-series data by direct density-ratio estimation., in: SDM, Vol. 9, SIAM, 2009, pp. 389–400.

[13] M. M. Salvador, B. Gabrys, I. Zliobaite, Online detection of shutdown periods in chemical plants: A case study, Procedia Computer Science 35 (2014) 580 – 588.

[14] S. Jimeng, F. Christos, P. Spiros, S. Y. Philip, Graphscope: parameter-free mining of large time-evolving graphs., in: Proceedings of the 13th

ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2007, pp. 687–696.

[15] H. Wang, M. Tang, Y. Park, C. E. Priebe, Locality statistics for anomaly detection in time series of graphs, IEEE Transactions on Signal Processing 62 (3) (2014) 703–717.

[16] A. Z. Jacobs, A. Clauset, A unified view of generative models for networks: models, methods, opportunities, and challenges, arXiv:1411.4070.

[17] S. Fortunato, Community detection in graphs, Physics reports 486 (3) (2010) 75–174.

[18] D. Jin, B. Gabrys, J. Dang, Combined node and link partitions method for finding overlapping communities in complex networks, Scientific Reports 5 (2015) 8600. doi:10.1038/srep08600.

[19] W. Wang, P. Jiao, D. He, D. Jin, L. Pan, B. Gabrys, Autonomous overlapping community detection in temporal networks: A dynamic bayesian nonnegative matrix factorization approach, Knowledge-Based Systems 110 (2016) 121 – 134. doi:https://doi.org/10.1016/j.knosys.2016.07.021.
URL http://www.sciencedirect.com/science/article/pii/S0950705116302350

[20] J. Xie, M. Chen, B. K. Szymanski, Labelrankt: Incremental community detection in dynamic networks via label propagation, in: Proceedings of the Workshop on Dynamic Networks Management and Mining, ACM, 2013, pp. 25–32.

[21] D. Greene, D. Doyle, P. Cunningham, Tracking the evolution of communities in dynamic social networks, in: Advances in social networks analysis and mining (ASONAM), 2010 international conference on, IEEE, 2010, pp. 176–183.

[22] P. Bródka, S. Saganowski, P. Kazienko, Ged: the method for group evolution discovery in social networks, Social Network Analysis and Mining 3 (1) (2013) 1–14.

[23] M. E. Newman, Modularity and community structure in networks, Proceedings of the national academy of sciences 103 (23) (2006) 8577–8582.

[24] M. Chen, T. Nguyen, B. K. Szymanski, A new metric for quality of network community structure, arXiv preprint arXiv:1507.04308.

[25] B. Karrer, M. E. Newman, Stochastic blockmodels and community structure in networks, Physical Review E 83 (1) (2011) 016107.

[26] C. E. Priebe, J. M. Conroy, D. J. Marchette, Y. Park, Scan statistics on enron graphs, Computational & Mathematical Organization Theory 11 (3) (2005) 229–247.

[27] T. P. Peixoto, Model selection and hypothesis testing for large-scale network models with overlapping groups, Physical Review X 5 (1) (2015) 011033.

[28] X. Yan, C. Shalizi, J. E. Jensen, F. Krzakala, C. Moore, L. Zdeborová, P. Zhang, Y. Zhu, Model selection for degree-corrected block models, Journal of Statistical Mechanics: Theory and Experiment 2014 (5) (2014) P05007.

[29] K. Henderson, B. Gallagher, L. Li, L. Akoglu, T. Eliassi-Rad, H. Tong, C. Faloutsos, It's who you know: graph mining using recursive structural features, in: Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, 2011, pp. 663–671.

[30] A. Clauset, C. Moore, M. E. Newman, Structural inference of hierarchies in networks, in: Statistical network analysis: models, issues, and new directions, Springer, 2007, pp. 1–13.

[31] A. Clauset, C. Moore, M. Newman, Hierarchical structure and the prediction of missing links in networks, Nature 453 (7191) (2008) 98–101.

[32] D. Liben-Nowell, J. Kleinberg, The link-prediction problem for social networks, Journal of the American society for information science and technology 58 (7) (2007) 1019–1031.

[33] I. Guyon, U. Von Luxburg, R. C. Williamson, Clustering: Science or art, in: NIPS 2009 workshop on clustering theory, 2009, pp. 1–11.

[34] M. Newman, Networks: an introduction, Oxford university press, 2010.

[35] M. E. Newman, The structure and function of complex networks, SIAM review 45 (2) (2003) 167–256.

[36] T. P. Peixoto, The graph-tool python library, figshare.

[37] A. Clauset, Network analysis and modeling (2013).

[38] N. Eagle, A. S. Pentland, Reality mining: sensing complex social systems, Personal and ubiquitous computing 10 (4) (2006) 255–268.

[39] B. Klimt, Y. Yang, The enron corpus: A new dataset for email classification research, in: European Conference on Machine Learning, Springer, 2004, pp. 217–226.

[40] R. Marks, Enron timeline, http://www.agsm.edu.au/bobm/teaching/be/enron/timeline.html.

[41] C. Erdman, J. W. Emerson, et al., bcp: an r package for performing a bayesian analysis of change point problems, Journal of Statistical Software 23 (3) (2007) 1–13.

[42] S. Aminikhanghahi, D. J. Cook, A survey of methods for time series change point detection, Knowledge and Information Systems (2016) 1–29.

[43] G. Kossinets, D. J. Watts, Empirical analysis of an evolving social network, science 311 (5757) (2006) 88–90.