Anomaly Detection in Time Series of Graphs using ARMA Processes

Brandon Pincombe[a]

Abstract

There are many situations in which indicators of changes or anomalies in communication networks can be helpful, e.g. in the identification of faults. A dynamic communication network is characterised as a series of graphs with vertices representing IP addresses and edges representing information exchange between these entities weighted by packets sent. Ten graph distance metrics are used to create time series of network changes by sequentially comparing graphs from adjacent periods. These time series are individually modelled as univariate autoregressive moving average (ARMA) processes. Each time series is assessed on the ability of the best ARMA model of it to identify anomalies through residual thresholding.

Introduction

Automatic highlighting of anomalously large changes in communications networks can be of benefit, e.g. as a tip off for fault detection or as an indicator of a change in the structure or use of a network. Computer communications networks can be characterised as graphs with vertices representing IP addresses and edges representing information exchange between these entities weighted by packet traffic. The similarity between a network graph for one day and that for another day can be represented using a variety of graph distance measures. Sequential comparison of the graph for each day with those for the previous day or days is used to produce a time series of graph distances using ten graph distance metrics. In order to determine what changes are abnormal an ARMA model of the normal changes in each of the time series is built and residuals exceeding a threshold are defined as anomalous.

This method is tested on a dataset of 102 days of TCP/IP traffic collected from 5 probes on an enterprise network excluding weekends and public holidays. The network administrators identified three days (22, 64 and 89) on which they thought the network had changed or behaved aberrantly. Only on day 64 was there a suggested reason: the introduction of a web based personnel management system (WBPMS). There may have been other network disruptions in this period but none were notable enough to be mentioned by network administrators.

If anomalies are defined as days with residuals of more than two standard errors from the best ARMA model the edit, MCS vertex, MCS edge, weight and weight common (MCS weight) distance metrics yield detections of WBPMS introduction with varying levels of false alarms. For the entropy and median distance metrics, detection occurs with a threshold just below two standard errors. The spectral, modality and diameter distance metrics do not allow ARMA detection of WBPMS introduction.

A similar pattern is seen for detections of all three anomalies with spectral, modality and diameter distance metrics not producing any detections, MCS edge, MCS vertex, edit, median and entropy metrics able to produce detections with no false alarms, and weight and MCS weight able to detect all anomalies with a small number of false alarms. Of the three anomalies the introduction of WBPMS was the most difficult to detect using ARMA based methods. In contrast it was the most easily detected using a comparison technique based on median edit graphs.

These results are promising and imply that high precision and recall ARMA based anomaly detection is possible when appropriate graph distance metrics are used to build a time series of network graph distances. Rigorous testing of whether a practical anomaly detection system can be constructed in this way can only be achieved by repeating this procedure on simulated time series of network graphs with anomalies of known type and magnitude injected into particular graphs.

[a] Intelligence, Surveillance and Reconnaissance Division, Defence Science and Technology Organisation, PO Box 1500, Edinburgh SA 5111, Australia (Email: Brandon.Pincombe@dsto.defence.gov.au)

From Distance Metrics to Time Series

The communications network for each period is characterised as a graph $G = (V,E)$ containing a finite set of vertices V and edges E. The vertices represent the communications nodes such as IP addresses, telephone numbers, etc. and the edges represent communications between these nodes. The edges are weighted by the volume of traffic along them. A number of graph topology distance measures are used to quantify the differences between the graph representations of the communication network. For each of these graph topology distance measures a time series of changes is constructed by comparing the graph for a given period with the graph(s) from one or more previous periods.

Metric	Vertices used?	Edges used?	Vertex weights used?	Edge weights used?	Range	Value if graphs identical
Weight	No	Yes	No	Yes	[0,1]	0
MCS Weight	No	Yes	No	Yes	[0,1]	0
MCS Edge	No	Yes	No	No	[0,1]	0
MCS Vertex	Yes	No	No	No	[0,1]	0
Graph Edit	Yes	Yes	No	No	[0,∞)	0
Median Edit	Yes	Yes	No	No	[0,∞)	0
Modality	No	Yes	No	Yes	[0,1]	0
Diameter	Yes	Yes	No	No	[0,∞)	0
Entropy	No	Yes	No	Yes	(-1,1)	0
Spectral	No	Yes	No	Yes	[0,1]	0

Table 1: Summary of metrics

Edges, $(u,v) \in E$, are defined by the pair of vertices, e.g. u and v, that they join and are directed if $(u,v) \in E$ is an ordered pair and undirected if it is not. Two vertices $u,v \in V$ are considered to be adjacent, $u \rightarrow v$, if there is an edge defined in terms of u and v.

Vertices, edges and their combinations associated with a graph, G, are referred to as elements. The domain of a weight function can be limited to edge elements, in which case it is called an edge-weight, or the vertex elements, called a vertex-weight, or span all edges and vertices, referred to as a total-weight. The weight values, w_V and w_E, assigned to elements of the graph $G=(V,E, w_V, w_E)$ are symbolic for vertex-weights, i.e. $w_V : V \rightarrow L_V$ where L_V are unique one-to-one labels for each $v \in V$, and numerical for edge-weights with the weight $w_E : E \rightarrow \Re^+$. All graphs have a unique one-to-one symbolic value for each vertex-weight and are thus considered to be labelled. The number of vertices in $G=(V,E)$ is denoted by $|V|$ and the number of edges by $|E|$.

All distance measures used are metrics. This means that the distance between two graphs is a positive real number, i.e. $d(G,H) \in \Re^+$, the zero distance is equivalent to graph isomorphism, all distance measures are symmetric, i.e. $d(G,H) = d(H,G)$, and they satisfy the triangle inequality, i.e. $d(G,F) \le d(G,H) + d(H,F)$.

The following graph topology distance measures rely on identification and comparison of the elements in common between graphs by finding maximum common subgraphs (Shoubridge et al., 1999). A subgraph of $G=(V_G, E_G, w_G^V, w_G^E)$ is a graph $S=(V_S, E_S, w_S^V, w_S^E)$ where $V_S \subseteq V_G$ and $E_S \subseteq E_G \cap (V_S \times V_S)$. The vertex-weight w_S^V of S is w_G^V restricted to V_S and the edge-weight w_S^E of S is w_G^E restricted to E_S. The maximum common subgraph (MCS) F of G and H, $F = \text{mcs}(G,H)$, is the common subgraph with the most vertices, i.e. no other common subgraph K of G and H exists, with more vertices than F.

The metrics are summarised in Table (1). It is worth noting that while vertex weights are not directly used in any of the measures vertex labels are used in all of them in order to make the process of comparing graphs easier and more accurate. Some graph theoreticians would consider these to be vertex weights. All metrics take the value of zero if the two graphs compared are identical, although in the case of median edit distance this means that the next graph in the series is identical to the median graph of the last five graphs. While several of the metrics are theoretically unbounded it is worth noting that edit distance produces integer results and that only diameter distance produces results that are far from zero in this case.

Weight Distance

This measure of graph distance sums the differences in edge-weight values over all edges in the two graphs, normalises this by the larger of the sums of the edge-weight values within each graphs and divides by the total number of edges in the double summation. One has:

$$d(G,H)=\frac{\sum_{u,v\in V}\frac{\left|w_E^G(u,v)-w_E^H(u,v)\right|}{\max\left\{w_E^G(u,v),w_E^H(u,v)\right\}}}{\left|E_G\cup E_H\right|},\ldots(1)$$

where $w_E^{(\bullet)}(u,v)$ is the weight of the edge joining u and v; and $d(G,H)$ is the distance between graphs G and H (Shoubridge et al., 1999).

MCS Weight Distance

This measure of graph distance strongly resembles the weight distance measure in Equation (1) but considers only those edges that appear in the maximum common subgraph (MCS).

MCS Edge Distance

The MCS edge distance metric is calculated by counting the number of edges in the MCS of two graphs, normalising this by the number of edges in the larger of the two graphs, and subtracting the result of this from one. The distance will always be in the interval [*0*,*1*] and the closer the distance is to *0* the more similar the graphs are. So

$$d(G,H)=1-\frac{\left|mcs(E_G,E_H)\right|}{\max\{|E_G|,|E_H|\}}, \quad \ldots(2)$$

where $mcs(E_G,E_H)$ is the number of edges in the maximum common subgraph of G and H and $max\ \{|E_G|,|E_H|\}$ is the maximum of the number of edges in either G or H (Shoubridge et al., 1999).

MCS Vertex Distance

The MCS vertex distance metric is calculated by counting the number of vertices in the MCS of two graphs, normalising this by the number of vertices in the larger of the two graphs, and subtracting the result of this from one. As with the MCS edge metric, the distance will always be in the interval [*0*,*1*], and the closer the distance is to *0* the more similar the graphs are. One has:

$$d(G,H)=1-\frac{\left|mcs(V_G,V_H)\right|}{\max\{|V_G|,|V_H|\}}, \quad \ldots(3)$$

where $mcs(V_G,V_H)$ is the number of vertices in the maximum common subgraph of G and H and $max\ \{|V_G|,|V_H|\}$ is the maximum of the number of vertices in either G or H (Shoubridge et al., 1999).

Graph Edit Distance

The graph edit distance between graphs G and H is calculated by evaluating the sequence of edit operations required to make graph G isomorphic to graph H using the formula

$$d(G,H)=|V_G|+|V_H|-2|V_G\cap V_H|+|E_G|+|E_H|-2|E_G\cap E_H|, \quad \ldots(4)$$

where E_G and V_G are the edges and vertices of graph G, and E_H and V_H are the edges and vertices of graph H (Sanfeliu and Fu, 1983; Messmer and Bunke, 1998; Dickinson et al., 2002). The computational complexity of this measure can be reduced by assuming unique labeling of the nodes in the graph (Dickinson et al., 2004).

Median Graph Edit Distance

The (set) median graph $\overline{G}$ of a sequence of n uniquely labeled graphs $S=(G_1,\ldots,G_n)$ minimises the sum of distances between itself and the members of S for a particular distance metric. The set median graph can vary depending on the distance metric, $d(G_i,G_j)$, chosen but the general formula is

$$\overline{G}=\underset{G\in S}{\arg\min}\sum_{i=1}^{n}d(G,G_i)\,. \quad \ldots(5)$$

Following Dickinson et al. (2002; 2001), the graph edit distance metric, described in Equation (4), is used both to construct $\overline{G}$ and to calculate the distance from $\overline{G}$ to other graphs. The set median graph $\tilde{G}_n$ is calculated from a sequence of uniquely labeled graphs ($G_{n-L+1}, \ldots, G_n$) in window of length *L*. This window length is arbitrarily chosen to be five in accordance with Dickinson et al. (2002; 2001).

The distance between $\tilde{G}_n$ and G_{n+1} is classified as abnormal if

$$d(\tilde{G}_n, G_{n+1}) \geq \alpha\varphi, \qquad \ldots(6)$$

where α is a parameter and φ is the average deviation of the graphs in the window, ($G_{n-L+1}, \ldots, G_n$), from the median graph, $\tilde{G}_n$, given by the equation

$$\phi = \frac{1}{L} \sum_{i=n-L+1}^{n} d\left(\tilde{G}_n, G_i\right). \qquad \ldots(7)$$

Modality Distance

The Modality distance between graphs *G* and *H* is the absolute value of the difference between the Perron vectors of these graphs. Algebraically this can be written as

$$d(G,H) = \| \pi(G) - \pi(H) \| \qquad \ldots(8)$$

where π (*G*) and π (*H*) are the Perron vectors of graphs *G* and *H* respectively. The Perron vector π_{mx1} satisifies the equation

$$A\pi = \rho\pi, \ \pi > 0, \ \sum_{i=1}^{m} \pi_i = 0 \qquad \ldots(9)$$

where A_{mxm} is the non-negative irreducible adjacency matrix with spectral radius ρ.

Diameter Distance

The Diameter distance between graphs *G* and *H* is the difference in the average longest shortest paths for each graph:

$$d(G,H) = \left| \sum_{v \in V_H} \max d(H,v) - \sum_{v \in V_G} \max d(G,v) \right|, \ldots(10)$$

where *maxd*(*G*,*v*) is the distance to the vertex in *G* farthest away from *v*, via the shortest path.

Entropy Distance

The "Entropy" distance between graphs *G* and *H* is defined using entropy-like measures associated with the corresponding graphs. One has:

$$d(G,H) = E(H) - E(G), \qquad \ldots(11a)$$

where $E(*)$ is the entropy-like measure of the edges:

$$E(*) = -\sum_{e \in E_*} \left(\tilde{w}_*^e - \ln \tilde{w}_*^e \right), \qquad \ldots(11b)$$

where

$$\tilde{w}_*^e = \frac{w_*^e}{\sum_{e \in E_*} w_*^e}, \qquad \ldots(11c)$$

is the normalized weight for edge *e*.

Spectral Distance

The spectral distance between graphs *G* and *H* is calculated by using the *k* largest positive eigenvalues of the Laplacian, so

$$d(G,H) = \sqrt{\frac{\sum_{i=1}^{k} (\lambda_i - \mu_i)^2}{\min\left\{ \sum_{i=1}^{k} \lambda_i^2, \sum_{i=1}^{k} \mu_i^2 \right\}}}, \qquad \ldots(12)$$

where λ_i represents the eigenvalues of the Laplace matrix for graph *G*, and μ_i represents the eigenvalues of the Laplace matrix for graph *H*.

ARMA Modelling

The Box-Jenkins (Box and Jenkins, 1976) approach to time series model building is followed. This formalises the finding of an ARIMA model that adequately represents the underlying process that originally generated the time series. The *integrated* step in ARIMA is unlikely to be necessary as time series of graph distance measures are likely to be stationary due to the nature of their generation. It is still critically important to check that they are stationary and that an ARMA model is sufficient to represent the underlying process that generated them. The three steps of Box-Jenkins modelling are identification, estimation and diagnostic checking.

Time series are tested for stationarity. Autocorrelation and partial autocorrelation functions are then examined to determine whether an autoregressive (AR), moving average (MA) or ARMA model is needed. The parameters of the process are estimated using least squares for AR processes and non-linear optimisation for MA processes.

Metric	ARMA parameters for optimal AIC	Calculation time (secs)	True pos.	False pos.	False neg.
Weight	AR(1) MA(1) MA(2)	0.42	3	3	0
MCS Weight	AR(1) AR(2) MA(1) MA(2)	0.18	3	2	0
MCS Edge	AR(1) MA(1) MA(2)	0.17	3	0	0
MCS Vertex	AR(1) MA(1) MA(2)	0.17	3	0	0
Graph Edit	AR(1) MA(1) MA(2)	0.16	3	0	0
Median Edit	AR(1) AR(2) MA(1)	1.17	3	0	0
Modality	AR(1) AR(2) MA(1) MA(2)	4.94	0	0	3
Diameter	AR(1) MA(1) MA(2)	2.96	1	3	2
Entropy	MA(1)	0.47	3	0	0
Spectral	AR(1) AR(2) MA(1) MA(2)	4.32	0	4	3

Table 2: Summary of results.

The model adequacy is then checked using Akaike's Information Criterion (AIC), Schwartz’s Bayesian Information Criterion (BIC) and the Durbin-Watson test statistic (Akaike, 1973; Schwarz, 1978; Johnston and DiNardo, 1997).

Results

The results for a comparison technique and for ARMA models of the time series constructed using the ten graph distance metrics are displayed below. In each of the graphs a threshold of twice the standard error is shown and, using this threshold, accurate detections are displayed as closed boxes, false alarms as open circles and false negatives as open boxes.

A results summary is given in Table (2). In this table the ARMA parameters used to generate the combination of true positives, false positives and false negatives shown in the rightmost three columns are chosen by selecting the model with the lowest AIC. The calculation times for the modality and spectral distances are particularly great due to these metrics calculating the eigenvalues and eigenvectors of the edge adjecency matrix.

Median Graph Edit Distance (non-ARMA)

The edit distance from a median edit graph (Dickinson et al., 2002a) has been used as a comparison method for detecting anomalous days. Through varying α the threshold defined as anomalous is altered. With an α of two the three anomalies are detected with thirteen false alarms. If α is set to three only the three anomalies are detected. At an α of four only two anomalies are detected and this continues until α=4.45. When α=4.46 only the introduction of WBPMS is considered to be anomalous and by α=4.47 no anomalies are detected. Unlike the ARMA based methods in the rest of the results section the median edit graph technique picks the introduction of WBPMS as the strongest, rather than the weakest, anomaly.

Weight Distance

The Dickey-Fuller unit root test statistic (-5.718<-3.497), the augmented Dickey-Fuller unit root test statistic with two lags (-4.340<-3.498) and the Phillips-Perron unit root test with two lags (-5.711<-3.4965) all indicate stationarity at the 1% level. There are significant autocorrelations of 0.495, 0.236 and 0.259 at lags one, 25 and 26 and significant partial autocorrelations of 0.495, -0.272 and 0.236 at lags one, two and 24. Neither the autocorrelations nor the partial autocorrelations drop away to zero so it appears unlikely that purely autoregressive or moving average processes will produce models with the lowest AIC or BIC. An AR(1) MA(1) MA(2) process produces the best AIC and BIC, and a good Durbin-Watson test statistic (1.984). The standard error is 0.0918 and, with a decision threshold of two standard errors, all anomalies are detected with three false alarms. It is possible to set a threshold to detect all three anomalies with only one false alarm. The introduction of WBPMS is returned as the least anomalous of the anomalous periods.

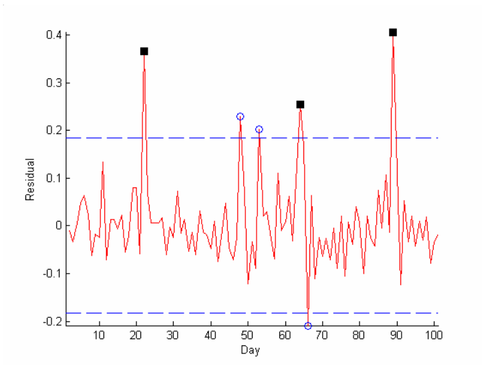

Figure 1. Residuals for weight distance.

MCS Weight Distance

The Dickey-Fuller unit root test statistic (-3.762<-3.497), the augmented Dickey-Fuller unit root test statistic with two lags (-4.897<-3.498), and the Phillips-Perron unit root test statistic (-5.872<-3.497), all indicate stationarity at the 1% level. There are significant autocorrelations of 0.442, 0.254 and 0.214 at one, 25 and 26 lags and significant partial autocorrelations of 0.442, -0.351 and 0.222 at lags of one, two and 24. The model with the best AIC and BIC was an AR(1) AR(2) MA(1) MA(2) process with an excellent Durbin-Watson test statistic (2.011). Its standard error was 0.0766 and use of a two standard error decision threshold detected all three anomalies with five false alarms. By setting at threshold above 0.170 but below 0.188, the three anomalies can be detected with two false alarms. Again, the introduction of WBPMS is considered the least abnormal of the anomalies.

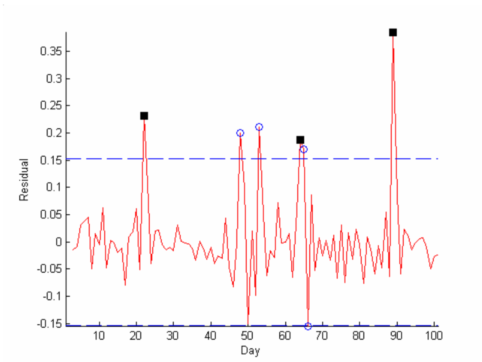

Figure 2. Residuals for MCS weight distance.

MCS Edge Distance

Using two lags, the augmented Dickey-Fuller test statistic (-4.250<-3.498) and the Phillips-Perron test statistic (-5.697<-3.497) both indicate stationarity at the 1% level. In the correlogram there are three significant autocorrelations of magnitude 0.496, 0.302 and 0.203 at lags one, 25, and 26 as well as two significant partial autocorrelations of magnitude 0.496 and -0.205 at lags one and two. The best AIC and BIC values for low order processes occurs for an AR(1) MA(1) MA(2) model that has a good Durbin-Watson test statistic (1.997). The standard error is 0.1459 and a threshold of two standard errors detects all three anomalies with no false alarms. WBPMS introduction returns the lowest residual of the anomalies.

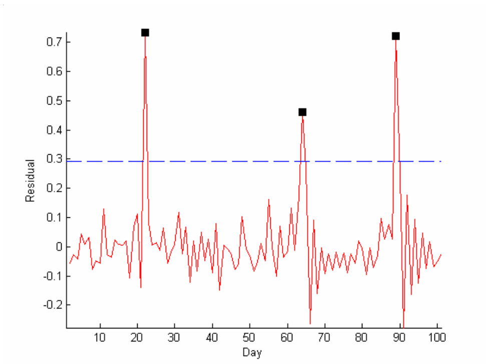

Figure 3. Residuals for MCS edge distance.

MCS Vertex Distance

For two lags, the augmented Dickey-Fuller test statistic (-4.355<-3.498) and the Phillips-Perron test statistic (-5.786<-3.497) indicate stationarity at the 1% level. The correlogram reveals three significant autocorrelations of magnitude 0.483, 0.265 and 0.201 at lags one, 25, and 26 as well as one significant partial autocorrelation of magnitude 0.483 at one lag. Oddly, the best AIC and BIC of low order processes is for an AR(1) MA(1) MA(2) model which also has by far the best Durbin-Watson statistic (2.047). The standard error is 0.1026 and a decision threshold of two standard errors detects all three anomalies with no false alarms. Introduction of WBPMS is returned as the least anomalous anomaly.

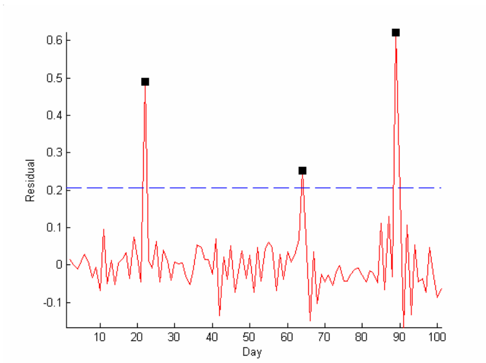

Figure 4. Residuals for MCS vertex distance.

Graph Edit Distance

The Dickey-Fuller unit root test statistic (-5.112<-3.497), the augmented Dickey-Fuller unit root test statistic with two lags (-3.593<-3.498) and the Phillips-Perron test statistic with two lags (-5.049<-3.497) all lie below the critical values indicating stationarity. Correlation analysis shows three significant autocorrelations of magnitude 0.576, 0.276 and 0.241 at lags one, two and three along with one partial autocorrelation of magnitude 0.576 at one lag. Both the autocorrelations and partial autocorrelations appear to drop away to zero but the partial autocorrelations do so more rapidly indicating that an AR process is more likely to be the best model of this time series than is an MA process. A standard set of low order models was tried on this time series and the lowest AIC and BIC occurred for an AR(1) MA(1) MA(2) model with an acceptable Durbin-Watson statistic (1.940). The standard error is 110.1 and a decision threshold of two standard errors detects all three identified anomalies with one false alarm. Were the threshold set between 264.4 and 383.6 all three anomalies would be detected with no false alarms.

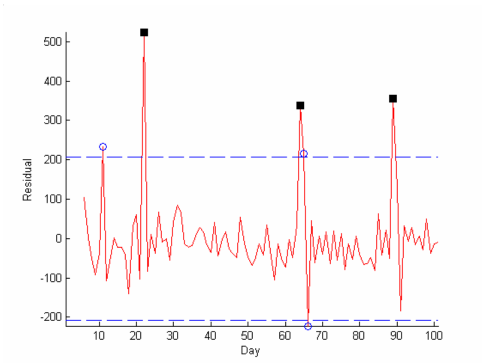

Figure 5. Residuals for edit distance.

Median Graph Edit Distance (ARMA)

Using two lags, the augmented Dickey-Fuller test statistic (-3.734<-3.4972) and the Phillips-Perron test statistic (-4.865<-3.497) both indicate stationarity at the 1% level. The correlogram reveals five significant autocorrelations of magnitude 0.589, 0.440, 0.286, 0.198 and 0.216 at lags one, two, three, six and 15 and a significant partial autocorrelation of 0.589 at a lag of one. Both the autocorrelations and the partial autocorrelations tail away to zero but this happens almost immediately for the partial autocorrelations suggesting that an autoregressive process may be the best model of this time series. The lowest AIC occurs for an AR(1) AR(2) AR(3) AR(6) AR(15) process but stationarity is not assured for 15 lags. An AR(1) AR(2) MA(1) process produces the best AIC and BIC for process with lags of low enough order to ensure stationarity. The Durbin-Watson test statistic (2.133) is adequate and the standard error is 0.2675. With a decision threshold of two standard errors two anomalies (but not WBPMS introduction) are detected with no false alarms. It is possible to set a threshold between 0.374 and 0.500 that allows all three anomalies to be detected with no false alarms.

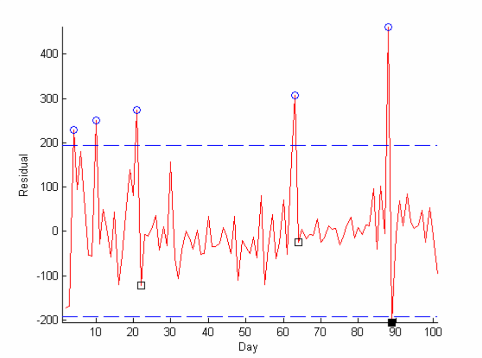

Figure 6. Residuals for median edit distance.

Modality Distance

The Dickey-Fuller test statistic (-9.948<-3.498), the doubly lagged augmented Dickey-Fuller test statistic and Phillips-Perron test statistic (-5.711<-3.497) all indicate stationarity. Correlation analysis shows a single autocorrelation of magnitude -0.232 at lag 26 and one partial autocorrelation of magnitude -0.220 at lag 26. As neither the autocorrelations nor the partial autocorrelations appear to drop away to zero, ARMA processes were tried but the lack of low lag correlations and autocorrelations implied it would be difficult to get an ARMA model to fit well. The best AIC and BIC were achieved for an AR(1) AR(2) MA(1) MA(2) process with a good Durbin-Watson statistic (2.039). The standard error was 0.5842 and no points were detected with a decision threshold of two standard errors. Almost all residuals were large and no level of threshold setting improved the performance without a very large number of false alarms.

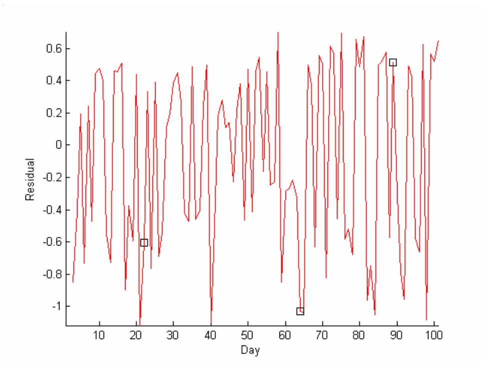

Figure 7. Residuals for modality distance.

Diameter Distance

The Dickey-Fuller unit root test statistic (-5.829<-3.497), the augmented Dickey-Fuller unit root test statistic with two lags (-3.760<-3.498) and the Phillips-Perron test statistic assuming two lags (-5.764< -3.497) indicate stationarity with a less than 1% chance of error. Correlation analysis reveals four significant autocorrelations of magnitude 0.484, 0.292, 0.274 and 0.264 at lags one, two, three and six as well as three significant partial autocorrelations of magnitude 0.484, 0.240 and -0.198 at lags one, six and eleven. Both the autocorrelations and partial autocorrelations appear to drop away but slowly indicating an ARMA model is likely to be best for this time series. The best AIC and BIC for low order processes for which the series is stationary occurs for an AR(1) MA(1) MA(2) model with an acceptable (1.947) Durbin-Watson test statistic. The standard error is 3452x10^3 and a decision threshold of two standard errors detects one anomaly and three false alarms. There is no threshold level returning the three anomalies without being swamped with false alarms. The WBPMS introduction is the least likely of the anomalies to be flagged for investigation as an anomaly.

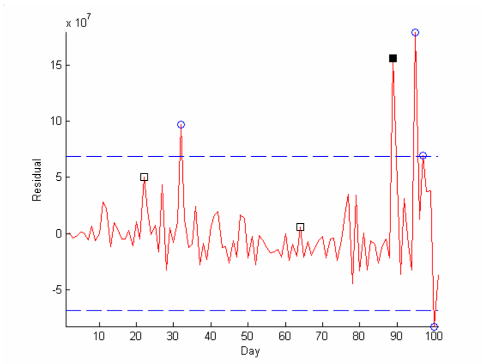

Figure 8. Residuals for diameter distance.

Entropy Distance

The Dickey-Fuller unit root test statistic (-14.60423 <-3.4972), the augmented Dickey-Fuller unit root test statistic with five lags (-6.821333 <-3.5000) and the Phillips-Perron unit root test with five lags (-21.36172<-3.4965) all lie below the critical values indicating stationarity. Correlation analysis shows two autocorrelations of magnitude -0.370 and -0.208 at lags one and two and three partial autocorrelation of magnitude -0.370, -0.400 and -0.220 at lags one, two and five. Both the autocorrelations and partial autocorrelations appear to drop away to zero but the autocorrelations do so more rapidly indicating that an MA process may be the best model of this time series. The minimum AIC and BIC values occur for an MA(1) model for which the Durbin-Watson test statistic of 1.794477 is acceptable. The standard error is 0.2646 and the use of two standard errors as a detection threshold detects two anomalous days (but not the WBPMS day) with no false alarms. It is possible to set a threshold greater than 0.420 but less than 0.490 allowing detection of all three anomalies without any false alarms.

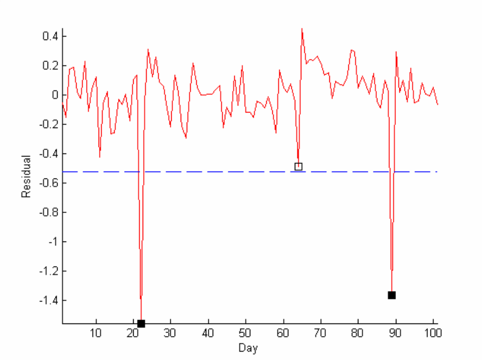

Figure 9. Residuals for entropy distance.

Spectral Distance

The Dickey-Fuller test statistic (-6.030<-3.497), augmented Dickey-Fuller test using four lags (-4.878<-3.498) and the Phillips-Perron test with four lags (-5.817<-3.497), allow rejection of the hypothesis of a unit root at the 1% level, therefore indicating stationarity. There are significant autocorrelations of 0.459, 0.240 and 0.207 at lags of one, five and 31 and significant partial autocorrelations of 0.459 and -0.211 at lags of one and two. The partial autocorrelations seem to drop away to zero so it is possible that a purely autoregressive processes will produce models with the lowest AIC. While an AR(1) model had the second lowest BIC, the lowest AIC and BIC occurred for an AR(1) AR(2) MA(1) MA(2) model with an acceptable Durbin-Watson statistic (2.169). The standard error was 0.5128 and using two standard errors as a decision threshold detected no anomalies but produced four false alarms. There is no threshold at which all three anomalies would have been detected without a very large number of false alarms.

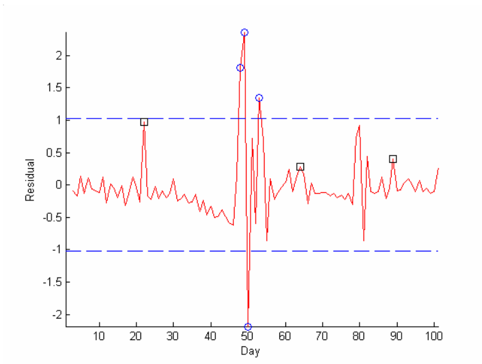

Figure 10. Residuals for spectral distance.

Conclusion

Time series based on the MCS edge, MCS vertex, edit, median and entropy metrics were all able to be modeled sufficiently well to detect all three anomalies with no false alarms. The weight and MCS weight metrics produced time series that could be modeled well enough to detect all three anomalies with one and two false alarms respectively. However, the time series based on the spectral, modality and diameter distance metrics did not lend themselves to accurate ARMA modelling. It is interesting to note that the spectral, modality and diameter distance metrics are considerably more computationally intensive than the others and give a much more finely detailed view of the distances between graphs.

The WBPMS introduction was the most difficult anomaly to detect using ARMA based methods. In contrast it was the most easily detected using a comparison technique based on median edit graphs. This problem may be associated with WBPMS introduction being a change point where the parameters of the best ARMA model of the underlying process may have been significantly altered. It is possible to find change points in time series through constructing multiple ARMA models for subsections of the series and testing the hypothesis that they are the same as each other. Such an approach will not detect anomalous days by itself, only change points. It also relies on having few enough change points to allow an adequate ARMA model to be built using the points between them, a condition met by this data set but not by all others.

These results are promising and imply that high precision and recall ARMA based anomaly detection is possible when appropriate graph distance metrics are used to build a time series of network graph distances. Ideally these graph distance metrics should be the MCS edge, MCS vertex, edit, median and entropy measures. Rigorous testing of whether a practical anomaly detection system can be constructed in this way can only be achieved by repeating this procedure on simulated time series of network graphs with anomalies of known type and magnitude injected into particular graphs.

Acknowledgements

Peter Dickinson and Miro Kraetzl provided the dataset and access to the Redback network analysis toolbox.

References

[1] H. Akaike, “Information theory and an extension of the maximum likelihood principle,” in B.N. Petrov and F. Csaki (eds.), “Second International Symposium on Information Theory” Budapest: Akademia Kiado, 267—281, 1973.

[2] G.E.P. Box and G.M. Jenkins, *Time Series Analysis: Forecasting and Control*, Revised Edition, Holden-Day, 1976.

[3] P. Dickinson, H. Bunke, A. Dadej and M. Kraetzl, “Median Graphs and Anomalous Change Detection in Communication Networks,” *Information, Decision and Control Conference (IDC-2002)*, Adelaide, Australia, pp. 59—64, 2002.

[4] P. Dickinson, H. Bunke, A. Dadej and M. Kraetzl, “Application of Median Graphs in Anomalous Change in Communications Networks,” *SCI2001/ISAS2001*, Orlando, Florida, 194—197, 2001.

[5] J. Johnston and J.E. DiNardo, *Econometric Methods*, 4th edition, McGraw-Hill, 1997.

[6] G. Schwarz, “Estimating the dimension of a model,” *Annals of Statistics*, 6, 461—464, 1978.

[7] P.J. Shoubridge, M. Kraetzl and D. Ray, “Detection of Abnormal Change in Dynamic Networks.” *Information, Decision and Control Conference (IDC'99)*, Adelaide, 557—562, 1999.