AUSTRALASIAN JOURNAL OF COMBINATORICS
Volume **35** (2006), Pages 299–311

Using graph diameter for change detection in dynamic networks

M. E. Gaston*

Department of Computer Science
University of Maryland Baltimore County
Baltimore, MD 21250
U.S.A.
mgasto1@cs.umbc.edu

M. Kraetzl

Intelligence, Surveillance and Reconnaissance Division
Defence Science and Technology Organisation
Edinburgh SA 5111
Australia
miro.kraetzl@dsto.defence.gov.au

W. D. Wallis

Department of Mathematics
Southern Illinois University
Carbondale, Illinois
U.S.A.
wdwallis@math.siu.edu

Abstract

Network dynamics has become a popular area of study as it is well known that networks evolve and adapt over time. With this in mind, abnormal change detection is critical to the understanding and control of network dynamics. This paper presents differences in graph diameter as a method for detecting abnormal change in a network time series. A formal definition of graph diameter is presented, with theoretical implications, examples and computational results. Also presented is an apparent means for characterization of network state without dependence on other networks in the time series. This leads directly to the ability to identify anomalous change and characterizing the affects on the network communications.

* On leave from US Department of Defense, Ft. Meade, MD 20755.

1 Introduction

We are primarily concerned with large-scale enterprise networks, the type of network which evolves from the communications of large military, business and social organizations. These are usually dynamic networks, and very often will be intranets (private internets, with restricted access). Often these enterprises can be represented by a time series of networks, each representing one day's or one week's transactions or communications.

Network management and other network analysis and control applications benefit greatly from effective and efficient means of detecting abnormal changes in such a time series of dynamic networks. Techniques for abnormal change detection are becoming more important as networks (especially large-scale enterprise networks) are becoming more complex and dynamic. (Of course, whether a particular abnormal change is significant for network management will depend on the particular network. This would have to be discussed on an *ad hoc* basis, and is beyound our scope here.)

We consider the analysis of large-scale complex networks on a macroscopic level, that is we consider properties of whole networks or large parts of networks as opposed to analyzing each individual network component, as would be more appropriate if one were dealing primarily with networks such as wireless networks (where local information is crucial to decision-making). Macroscopic analysis often provides an efficient means for network management and control. Existing techniques include various pattern recognition approaches [4, 7, 10, 11, 14, 15], spectral graph theory [13] and mean/median graphs [5, 8, 9]. An overview of the existing theory and applications can be found in [13, 12].

Because communications and other types of networks can be modeled as directed graphs (digraphs), the problem of abnormal change detection in networks becomes a problem of identifying meaningful graph measures that quantify the difference in a network time series. Graph diameter, a macroscopic measure of the "size" of a network, lends itself directly to abnormal change detection in a time series.

The paper contains a brief introduction to modeling communications networks as graphs in Section 2. Section 3 discusses the property of graph diameter and recent applications using graph diameter. After developing a distance measure based on graph diameter, Section 4 proves that the distance measure is a pseudo-metric. In Section 5, a sensitivity analysis is presented for the distance measure on some well known families of graphs. An example is presented in Section 6 and Section 7 contains an analysis of a real time series using difference in graph diameter. Section 8 proposes the use of graph diameter as a characterization of network state and concluding remarks are in Section 9.

2 Communications Networks as Graphs

Networks are ubiquitous. They are present in systems ranging from social networks to food networks, and from biological networks to communications networks. These networks can be represented by the mathematical abstraction known as graphs. A *graph* $G = (V, E)$ is a model of a network which consists of a finite vertex set V and a finite edge set E. An edge $e \in E$ is a pair of vertices denoting the endpoints of the edge. Edges in E can be directed or undirected. Directed edges are ordered pairs of vertices $(u, v) \in E$, with the edge oriented from the vertex u to the vertex v. An undirected edge is an unordered pair of vertices $u, v \in E$ denoting a connection between u and v. An undirected edge $u, v \in E$ can be modeled as directed by creating two directed edges (u, v) and (v, u). Graphs with directed edges are known as *directed graphs*, or *digraphs*. The number of vertices in $G = (V, E)$ is $|V|$ and the number of edges is $|E|$.

A *weight* function can also be introduced on the graph elements of G; it can be defined over the set of vertices or the set of edges. The *vertex-weight* and *edge-weight* are denoted by ω_V and ω_E respectively. Such a function maps the set of graph elements to associated attributes. These attributes could be vertex labelings, edge capacities, or traffic on an edge. A communications network can be represented as the graph $G = (V, E, \omega_V, \omega_E)$. Here, the vertex-weight, ω_V, is a labeling of the vertices, $\omega_V : V \to L$, where L is the set of values of some attribute of a given communications vertex (examples include cost, latency, etc.). Similarly, the edge-weight, ω_E, is a mapping of the edge to its measured traffic flow, $\omega_E : E \to \mathbb{R}$. In this paper we are not concerned with vertex-weights, although it is quite possible to use some measure of vertex-weights for change detection.

3 Graph Diameter

Graph diameter has become a popular property of graphs in recent research in the fields of complex interactive networks and network analysis. For error and attack tolerance of networks, graph diameter has been used to demonstrate change in networks subject to different faults [1]. Small-world networks are defined in terms of graph diameter and neighborhood size [6, 16, 17]. Tying the ideas together, the diameter of the World-Wide Web and the Internet have been measured, and are used as evidence in the discussion of these two networks as small-world networks [1, 2].

Graph diameter can take on several meanings. We shall be concerned with macroscopic structure of networks, so graph diameter will be defined based on averages of the longest shortest paths for all vertices. The definition of graph diameter more commonly used in graph theory is the greatest of the longest shortest paths for all vertices [18].

The *shortest path* between two vertices u and v is the path that starts at u and ends at v with the smallest possible number of vertices visited in between. The shortest path can also be based on an edge-weight, where the shortest path is the path with the smallest sum of weights for all edges traversed between vertices u and v. There are well known efficient algorithms for computing shortest paths [18].

Eccentricity is the key property that leads to the definition of graph diameter and its application to abnormal change detection.

Definition 3.1 *Given a graph $G = (V, E)$ the eccentricity of a vertex $v \in V$, denoted $\varepsilon(v)$, is the maximum distance from v to any other vertex in the graph. That is,*

$\varepsilon(v) = \max_{u \in V} d(v,u)$, where $d(v,u)$ is the length of the shortest path between the vertex v and u.

The shortest path used in the definition of eccentricity can either be based on number of hops (each edge weight equal to one), or on an edge-weight as described above. Graph diameter is defined based on the eccentricities of all vertices in the graph. Here, graph diameter is presented as the average of all vertex eccentricities.

Definition 3.2 *Given a graph $G = (V, E)$, graph diameter is defined as*

$$D(G) = \frac{\sum_{v \in V} \varepsilon(v)}{|V|}.$$

Of course, graph diameter can be weighted or not weighted. Graph diameter can be calculated using algorithms for all pairs of shortest paths. From the algorithms, averaging over all of the vertex eccentricities is straightforward. Section 6 presents an example network for calculating vertex eccentricities and graph diameter.

To use graph diameter as a measure of change, or difference, between two graphs G and H, simply calculate the difference between their respective diameters. The difference, $f(G, H)$, is given by

$$f(G, H) = |D(G) - D(H)|.$$

The absolute value is used to measure the magnitude of the difference in the diameters of the two graphs. The following two sections discuss some theoretical properties of this distance measure and some sensitivities that this measure has for detecting change in networks.

4 A Pseudo-Metric

It is a desirable property that a measure of network distance be a *metric*. A metric on set A is a function, $d : A \times A \to \mathbb{R}$, that maps a product set, $A \times A$, to the real numbers such that four axioms hold true: non-negativity, separation, symmetry, and the triangle inequality.

The axiom of separation states that for a given distance measure, $d(a, b) = 0$ if and only if $a = b$. In terms of graphs, two graphs are equivalent if they are *isomorphic*, that is, the two graphs have a one-to-one mapping of the vertices which preserve the edge connections [18]. It is certainly possible that two graphs, G and H, are not isomorphic, but have the same graph diameter and therefore $f(G, H) = 0$. A simple example of this is the complete graph with n vertices, K_n. The complete graph K_n has diameter $D(K_n) = 1, \forall\ n \geq 2$, and it follows that $f(K_i, K_j) = 0, \forall\ i, j \geq 2$; however, K_i and K_j are not isomorphic if $i \neq j$.

Although the function $f(G, H) = |D(G) - D(H)|$ is not a metric, it is still a useful measure of distance between two graphs. To prove this, we present the notion of a *pseudo-metric*.

Definition 4.1 *A pseudo-metric on set A is a function, $d : A \times A \to \mathbb{R}$, for which the following axioms hold for all $a, b, c \in A$:*

(i) $d(a,b) \geq 0$,

(ii) $d(a,b) = 0$, *if* $a = b$,

(iii) $d(a,b) = d(b,a)$,

(iv) $d(a,b) \leq d(a,c) + d(c,b)$.

For the graph diameter distance measure, axiom (ii) of Definition 4.1 implies that if two graphs are isomorphic, denoted $G \cong H$, then they have the same graph diameter, but having the same graph diameter does not imply that the two graphs are isomorphic. This relaxes the separation axiom in the definition of a metric.

Theorem 4.1 *The function $f(G,H) = |D(G) - D(H)|$ is a pseudo-metric.*

Proof. It must be shown that all axioms in Definition 4.1 hold for $f(G,H) = |D(G) - D(H)|$. For (i), by definition of absolute value, $|D(G) - D(H)| \geq 0, \forall\, G, H$, so $f(G,H) \geq 0$.

Next, to show that (ii) holds, the relation of graph isomorphism is used. If G and H are isomorphic, then there exists a bijection $g : V(G) \to V(H)$ such that $\{u,v\} \in E(G)$ implies that $\{g(u), g(v)\} \in E(H)$. Due to this bijection $|V(G)| = |V(H)|$ and $\varepsilon(v) = \varepsilon(g(v))$ and therefore $f(G,H) = 0$ if $G \cong H$.

Again, by definition of absolute value, $f(G,H) = |D(G) - D(H)| = |D(H) - D(G)| = f(H,G)$, so (iii) holds.

Finally, it must be shown that the triangle inquality $f(G,H) \leq f(G,K) + f(K,H)$ holds for all graphs G, H, K. Without loss of generality it is assumed that $D(G) \geq D(H)$. Three cases are considered. In the first case, let $D(G) \geq D(H) \geq D(K)$. Then

$$|D(G) - D(K)| + |D(K) - D(H)| = D(G) - D(K) + D(H) - D(K),$$

but $D(H) - D(K) \geq 0$ and $-D(K) \geq -D(H)$, so

$$\begin{aligned} D(G) - D(K) + D(H) - D(K) &\geq D(G) - D(K) + 0 \\ &\geq D(G) - D(H) \\ &= |D(G) - D(H)|. \end{aligned}$$

For the second case, let $D(G) \geq D(K) \geq D(H)$. Then

$$\begin{aligned} |D(G) - D(K)| + |D(K) - D(H)| &= D(G) - D(K) + D(K) - D(H) \\ &= D(G) - D(H) \\ &= |D(G) - D(H)|. \end{aligned}$$

In the last case, let $D(K) \geq D(G) \geq D(H)$. Then

$$|D(G) - D(K)| + |D(K) - D(H)| = D(K) - D(G) + D(K) - D(H),$$

but $D(K) - D(G) \geq 0$ and $D(K) \geq D(G)$, so

$$\begin{aligned} D(K) - D(G) + D(K) - D(H) &\geq 0 + D(K) - D(H) \\ &\geq D(G) - D(H) \\ &= |D(G) - D(H)|. \end{aligned}$$

Axioms (i)-(iv) hold, therefore $f(G,H) = |D(G) - D(H)|$ is a pseudo-metric. ■

It is important to note that it is possible to define an equivalence class for which the function f is a metric. If we define $\mathfrak{G}$ to be the set of all networks with diameter equal to $D(G)$ and write $D(\mathfrak{G})$ for the (common) diameter of members of $\mathfrak{G}$, then $f(\mathfrak{G}, \mathfrak{H}) = |D(\mathfrak{G}) - D(\mathfrak{H})|$ is a metric on the (quotient set of) equivalence classes $\{\mathfrak{G}\}$.

5 Sensitivity Analysis

In addition to being a pseudo-metric, if difference in graph diameter is to be a useful indicator of changes in network topology, it is important that small changes in a relatively dense unweighted network should not result in large changes in the diameter. To investigate the sensitivity of our distance measure, we look at two extreme cases, the families of complete graphs and of simple cycles. While real-world networks do not usually come close to such highly-structured examples, they make useful benchmarks.

The eccentricity of any vertex of the complete graph K_n is 1, so the diameter is $D(K_n) = 1$. If one edge is deleted, two vertices of the resulting graph G will have eccentricity 2, so $D(G) = \frac{n+2}{n}$. If k edges are deleted, where $k \leq \frac{n}{2}$, at most 2k vertices will have eccentricity 2, and the others will have eccentricity 1. So $D(G) \leq 2$.

In general, if a graph G has any vertex x of degree $n-1$, then $\varepsilon(x) = 1$. Any two other vertices y, z are joined by at least one path of length 2, namely yxz, so $\varepsilon(y) \leq 2$ and $\varepsilon(z) \leq 2$. So $D(G) < 2$. This means that even when $\frac{n}{2}$ edges are deleted from K_n, the resulting graph has diameter less than 2 unless the $\frac{n}{2}$ edges form a one-factor.

At the other extreme, consider the cycle C_n. Every vertex has eccentricity $\lfloor \frac{n}{2} \rfloor$, so this is the graph diameter. If C_n has vertices $x_1, x_2, \ldots, x_n$, let C_n^i denote the result of adding edge $x_n x_i$ (a *chord* of length i) to C_n. The addition of an edge reduces the graph diameter. As an example, we have

$$D(C_{10}) = 5, D(C_{10}^2) = 4.2, D(C_{10}^3) = 4.4, D(C_{10}^4) = 3.8, D(C_{10}^5) = 4.0,$$

while

$$D(C_9) = D(C_9^2) = 4, D(C_9^3) = D(C_9^4) = 3.56.$$

Theorem 5.1 *If* $n \equiv 0 \pmod 4$, *then*

$$D(C_n^i) \geq D(C_n^{\frac{n}{2}}) = \frac{3n}{8}.$$

If $n \equiv 2 \pmod 4$, *then*

$$D(C_n^i) \geq D(C_n^{\frac{n}{2}-1}) = \frac{3n}{8} + \frac{4}{8n}.$$

If $n \equiv 1 \pmod 4$, *then*

$$D(C_n^i) \geq D(C_n^{\lfloor\frac{n-1}{2}\rfloor}) = \frac{3n+1}{8}.$$

If $n \equiv 3 \pmod 4$, *then*

$$D(C_n^i) \geq D(C_n^{\lfloor\frac{n-3}{2}\rfloor}) = \frac{3n}{8} - \frac{14n-15}{8n}.$$

Proof. In the first case, it is easy to see that the graph diameter is smallest when the chord is of length $n/2$: in those cases the eccentricities are $n/2-1$ (twice — the vertices furthest from the endpoints of the chord), $n/2-2$ (four times — the vertices adjacent to the two just mentioned), $n/2-3$ (four times), ... , $n/4+1$ (four times), $n/4$ (twice — the endpoints of the chord). The average is as shown.

The other cases are handled similarly. It is interesting to note that when $n \equiv 2 (\mathrm{mod} 4)$, $D(C_n^{\frac{n}{2}}) = D(C_n^{\frac{n}{2}-1}) + \frac{n-2}{2n}$. ■

Theorem 5.1 shows that the addition of an edge can reduce the diameter of C_n by approximately 25%, from $\lfloor\frac{n}{2}\rfloor$ to approximately $\frac{3n}{8}$. This discussion shows that the distance measure based on graph diameter is sensitive to small changes in sparse graphs and less sensitive for small changes in near complete graphs.

This analysis is pointless in the case of weighted networks, because the difference in diameter due to inclusion or deletion of one edge can be made as large as we please by increasing the weight of the edge.

6 An Example Network

Here an example network (graph) is used to demonstrate the calculation of vertex eccentricities and difference in graph diameter as presented in Section 3. First, the graph without an edge-weight function will be used, then an edge-weight ω will be introduced.

Let G be the graph in Figure 1 with $V = \{1, 2, \dots, 7\}$ and $E = \{\{1,2\}, \{1,3\}, \{2,5\}, \{3,5\}, \{4,7\}, \{5,6\}, \{5,7\}, \{6,7\}\}$[1]. The eccentricities for the example graph are: $\varepsilon(1) = 4$, $\varepsilon(2) = 3$, $\varepsilon(3) = 3$, $\varepsilon(4) = 4$, $\varepsilon(5) = 2$, $\varepsilon(6) = 3$, $\varepsilon(7) = 3$. Taking an average of all the eccentricities, $D(G) \approx 3.143$.

To demonstrate that graph diameter can detect small changes in graph topology, an edge is added to the graph in Figure 1. Figure 2 shows the graph with the edge

[1]The example graph is undirected, therefore, the edges are just unordered pairs of adjacent vertices. It is trivial to convert the graph to directed as described in Section 2.

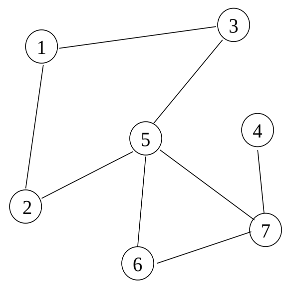

Figure 1: Example graph.

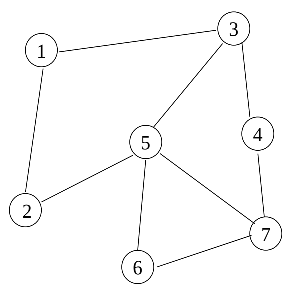

Figure 2: Example graph with edge {3,4} added.

$\{3,4\}$ added. Let this graph be $H = (E \cup \{3,4\}, V)$. It is left to the reader to verify that the graph diameter $D(H) \approx 2.714$ and the difference $f(G,H) \approx 0.429$.

Now, an edge-weight ω_1 is introduced on the original graph G to give $G_1 = (V, E, \omega_1)$. For simplicity, let $\omega_1(e) = 3, \forall e \in E$. This yields a graph diameter, $D(G_1) \approx 9.429$. To demonstrate that graph diameter can detect a change in edge-weight, set $\omega_2 = \omega_1, \forall e \neq \{3,5\} \in E$. Let $\omega_2(\{3,5\}) = 1$ and $G_2 = (E, V, \omega_2)$. With this minimal change, $D(G_2) = 8.0$, and $f(G_1, G_2) \approx 1.429$.

7 Time Series Using f

The examples in Section 6 demonstrate that graph diameter can be used to detect small changes in both network topology (unweighted) and network traffic (weighted). In this section, graph diameter is used to measure change in a time series of an actual communication network. Both unweighted and weighted time series are presented where the corresponding edge-weight is traffic. The time series are based on daily traffic activity.

We analysed sample network traffic which was used for previous investigations (see, for example, [3]). It was collected from the network management system of a large enterprise data network. The network can be thought of as an intranet with some 70,000 users. A traffic probe was installed on a single physical link in the network and traffic parameters were logged over 24 hour periods in daily log files. This probe collected traffic between up to 9,000 users daily. A traffic log file contains information on the logical originators (O) and destinations (D) of traffic (derived from network address information) and the volume of traffic transmitted between O-D pairs. To reduce the overall number of O-D pairs in the data set, the individual users (network addresses) were clustered by the business domain they belonged to on the data network. The aggregated logical flows of traffic between the 325 business domains observed over this physical link in a day were then represented as a directed and labelled graph. Vertex-weight identified the business domains of logical nodes communicating over the physical link with edge-weight denoting the total traffic transmitted between corresponding O-D pairs over a 24 hour period.

Successive log files collected over subsequent days produced a time series of corresponding directed and labelled graphs representing traffic flows between business domains communicating over the physical link in the network. Log files were collected continuously over a period of several months, from July 9 to December 24, 1999, and weekends and public holidays were removed to produce a final set of 102 log files representing the successive business days' traffic.

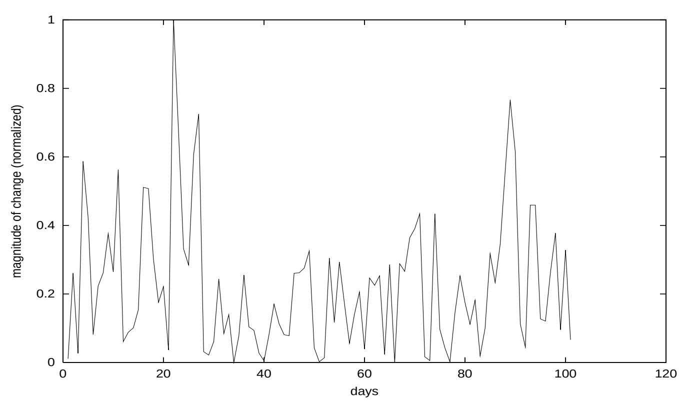

Figure 3: Change in time series, using unweighted comparison.

Figure 3 shows the times series change based on the difference in graph diameter normalized by the maximum change over the interval. Peaks in the time series represent potential abnormal change in network topology. Abnormal change would be based on deviations from some normal state of the network. Examples of network changes corresponding to the maximum peak in Figure 3 are given in Figures 4 and 5 respectively.

Figure 6 shows the time series change based on the difference in weighted graph diameter, normalized by the maximum change over the interval. As in the plot in Figure 3, peaks represent potential abnormal change. The peaks and valleys

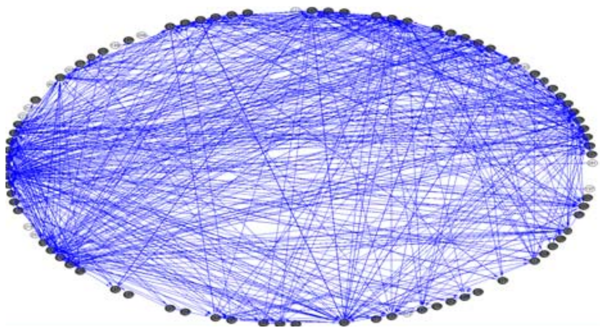

Figure 4: Enterprise network before change.

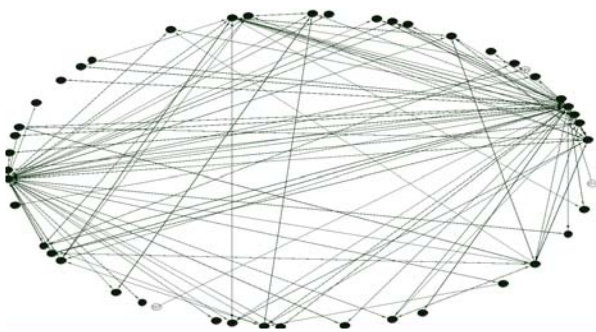

Figure 5: Enterprise network after change.

differ between the two plots, as different factors are compared for weighted and non-weighted graphs. The non-weighted plot would show topological changes where edges are either added or deleted. The weighted plot, based on edge traffic, can show both topological changes as well as abnormal changes in edge traffic. As demonstrated in Section 4 and Section 6, both are sensitive to small changes in the network.

8 Time Series Using D

Instead of considering time series of f (difference in graph diameters), one can use the diameter D directly and obtain an interesting time series of graphs. With the measure D, we simply use the characterization of a graph independent of any other graphs. This is the first time that we have addressed the problem of abnormal change detection from a characterization of an individual graph. Note that f (as a function of time) is effectively the rate of change of D in time.

As can be seen in Figure 7, there are still peaks and valleys in the time series plot of graph diameter, but these peaks and valleys have a different connotation than in

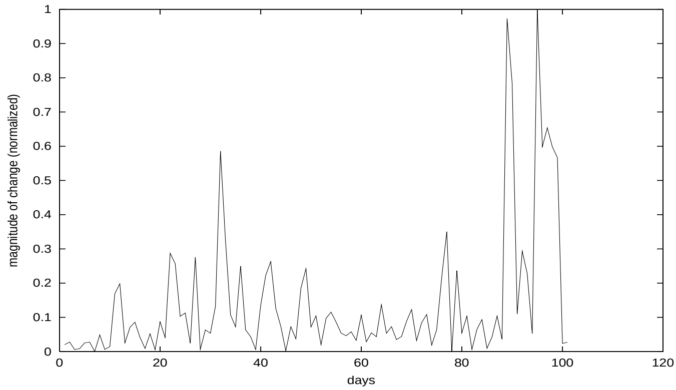

Figure 6: Change in time series, using weighted comparison.

the time series. This connotation is based on the difference in two graphs. From a communications network perspective, large graph diameter implies a longer average communication path between any two vertices. From the aspects of performance and redundancy, small graph diameter is desirable so as to minimize the longest average communication path link. So from this standpoint, an abnormal event could be defined as an event where the graph diameter moves above some established threshold. We also notice the obvious correlation of significant events with the time series using the consecutive f-distances (see Figure 3).

This is an interesting new approach to classifying change in time series of networks. Comparison with other networks is not necessary, but can still be used. The measure of graph diameter provides the ability to characterize a given network's state and we plan to pursue other aspects of this type of classification.

9 Conclusion

Graph diameter has been presented as an effective way of identifying change in dynamic networks. This novel approach to time series analysis has shown that both topological changes and traffic changes can be detected in communications networks. Graph diameter differs from other approaches in that it treats the network at a more macroscopic level and also allows for measuring the state of the network at any given time. Other approaches merely focus on the difference between two networks in series.

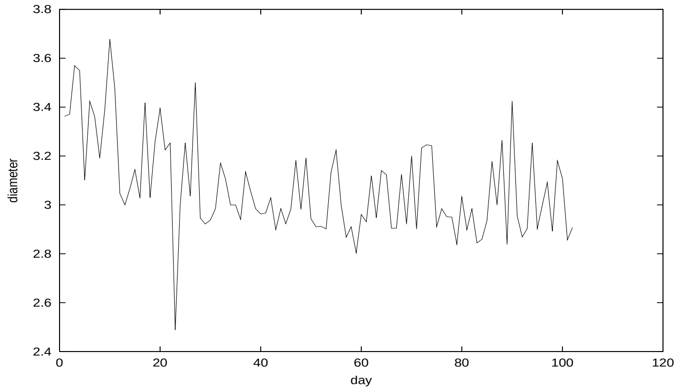

Figure 7: Characterization by diameter.

References

[1] R. Albert, H. Jeong and A. Barabási, The diameter of the world-wide web, *Nature* **401** (1999), 130–131.

[2] R. Albert, H. Jeong and A. Barabási, Error and attack tolerance of complex networks, *Nature* **406** (2000), 378–382.

[3] H. Bunke, M. Kraetzl, P. J. Shoubridge and W. D. Wallis, Measuring Change in Large Enterprise Data Networks, *Proc. Int. Conf. Information, Decision and Control* (IEEE, Adelaide, 2002), 53–58.

[4] H. Bunke and B. Messmer, Recent advances in graph matching, *Internat. J. Pattern Recognition Artif. Intell.* **11** (1997), 169–203.

[5] H. Bunke, A. Muenger and X. Jiang, Combinatorial search versus genetic algorithms: A case study based on the generalized mean graph problem, *Pattern Recognition Lett.* **20** (1999), 1271–1277.

[6] D. C. Callaway, M. E. J. Newman, S. H. Strogatz, and D. J. Watts. Network robustness and fragility: Percolation on random graphs, *Phys. Rev. Lett.* **85** (2000), 5468–5471. Available at `http://www.arxiv.org/abs/cond-mat/0007300`.

[7] D. G. Corneil and C. C. Gotlieb, An efficient algorithm for graph isomorphism, *J. ACM* **17** (1970), 51–64.

[8] X. Jiang and H. Bunke, On median graphs: Properties, algorithms, and applications, *IEEE Trans. Pattern Anal. Machine Intell.* **23** (2001), 1144–1151.

[9] X. Jiang, A. Muenger and H. Bunke, Synthesis of representative graphical symbols by computing generalized means, In *A. Chhabra, D. Dori (Eds.), Graphical Recognition, LNCS 1941* (Springer-Verlag, 2000), 183–192.

[10] B. T. Messmer and H. Bunke, A new algorithm for error-tolerant subgraph isomorphism detection, *IEEE Trans. Pattern Anal. & Machine Intell.* **20** (1998), 493–504.

[11] A. Sanfeliu and K. S. Fu, A distance measure between attributed relational graphs for pattern recognition, *IEEE Trans. Systems, Man. Cyber.* **13** (1983), 353–362.

[12] P. Shoubridge, M. Kraetzl and D. Ray, Detection of abnormal change in dynamic networks, In *Proc. Int. Conf. Information, Decision and Control*, IEEE, Adelaide, 1999, 557–562.

[13] P. J. Shoubridge, M. Kraetzl, W. D. Wallis and H. Bunke, Detection of abnormal change in a time series of graphs, *J. Interconnection Networks* **3** (2002), 85–101.

[14] J.R. Ullman, An algorithm for subgraph isomorphism, *J. ACM* **23** (1976), 31–42.

[15] W. D. Wallis, P. J. Shoubridge, M. Kraetzl and D. Ray. Graph distances using graph union, *Pattern Recognition Letters* **22** (2001), 701–704.

[16] D. J. Watts, *Small Worlds*, Princeton University Press, 1999.

[17] D. J. Watts and S. H. Strogatz, Collective dynamics of 'small-world' networks, *Nature*, **393** (1998), 440–442.

[18] D. B. West, *Introduction to Graph Theory* (Prentice Hall, New Jersey, 1996).

(Received 1 Mar 2005; revised 18 Aug 2005)