Computers & Industrial Engineering 144 (2020) 106451

Contents lists available at ScienceDirect

Computers & Industrial Engineering

journal homepage: www.elsevier.com/locate/caie

Monitoring binary networks for anomalous communication patterns based on the structural statistics

Panpan Zhou[a], Dennis K.J. Lin[b], Xiaoyue Niu[b], Zhen He[c,*]

[a] School of Management Science and Engineering, Nanjing University of Finance and Economics, Nanjing, China
[b] Department of Statistics, Pennsylvania State University, PA, USA
[c] Col lege of Management and Economics, Tianjin University, Tianjin, China

ARTICLE INFO

Keywords:
Network monitoring
Control charts
Exponential random graph models
Directed networks
Undirected networks

ABSTRACT

Network monitoring has become increasingly popular in the area of statistical process control due to its wide applications in fraud detection, corporate management and political behavioral analysis. This paper focuses on the cases where the communication pattern is significant while changes in specific nodes are negligible. The important features including the density, reciprocity, degree variability, and transitivity are considered to reflect the commonly-encountered communication patterns in social networks. The structural statistics are provided for characterizing the main features. A multivariate control chart is adopted to monitor the structural statistics simultaneously so as to account for their correlations and to decrease the overall false alarm rate. A performance evaluation framework is proposed based on the Exponential Random Graph Models (ERGMs) in order to simulate the shifts of communication patterns. The results of the numerical experiments show that the Hotelling T^2 control chart for the structural statistics outperforms several benchmark methods especially in detecting the large shifts of reciprocity and transitivity. The effectiveness of the proposed method is validated through the analysis of the Enron email communication networks.

1. Introduction

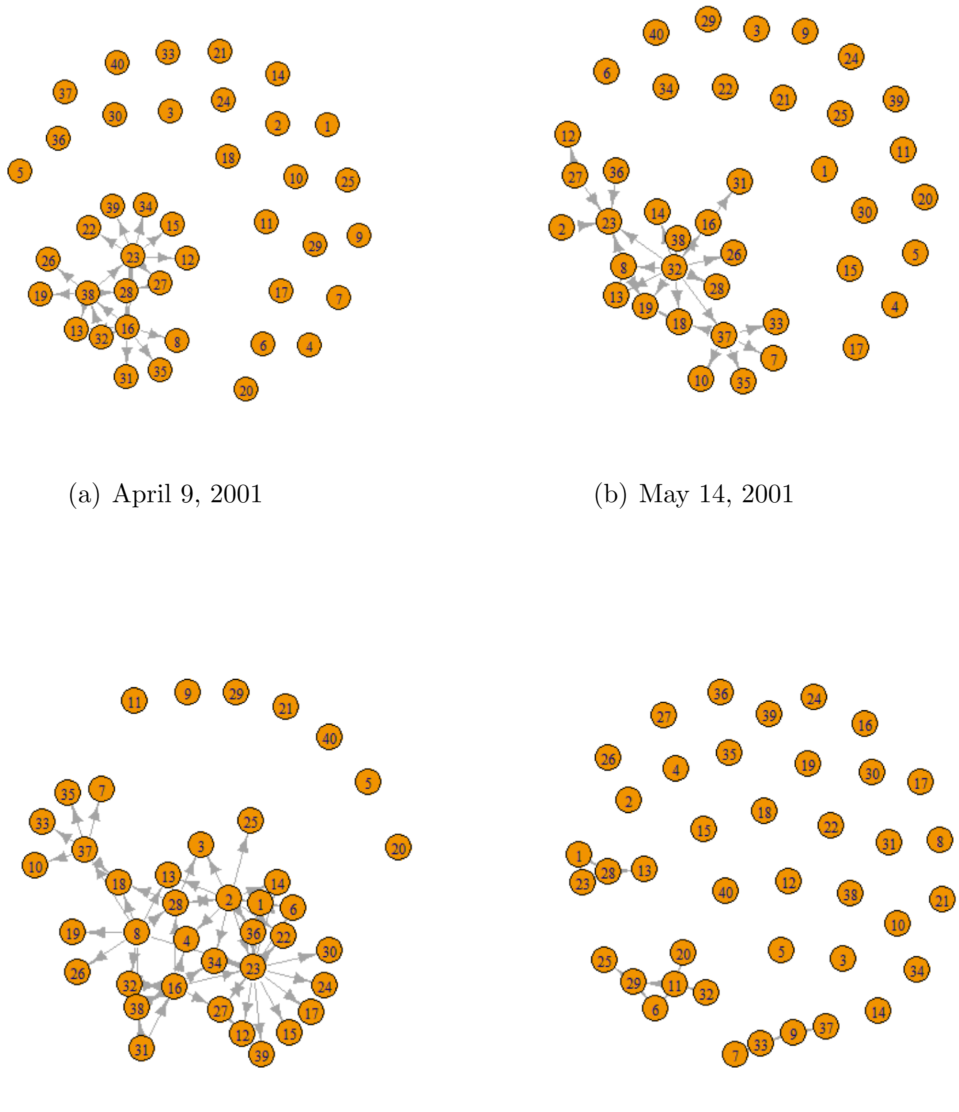
Networks describe the interactions among connected nodes. Network monitoring is to detect anomalous behaviors among nodes over time. It has wide applications in corporate governance, political behavioral analysis and Internet surveillance among others. A typical strategy for monitoring network data is to apply statistical process control techniques for monitoring network features. Usually the network features are characterized through structural statistics or statistical model (Savage, Zhang, Yu, Chou, & Wang, 2014; Woodall, Zhao, Paynabar, Sparks, & Wilson, 2017). For example, the density, the reciprocity, the degree variability, and the transitivity features are very commonly-used for describing the social relations. These features are interpreted as the overall frequency of node interactions, tendency of mutual interactions, node heterogeneity, and clustering effects of "friends of mine are my friends", respectively (Frank & Strauss, 1986; Holland & Leinhardt, 1981; Snijders, Pattison, Robins, & Handcock, 2006). In terms of structural statistics, the total numbers of local structures such as edges, mutual dyads, 2-stars, and triangles can be used to summarize these representative features. For statistical models, the probability of a network is modeled as an exponential function of these structural statistics, which is widely known as the Exponential family Random Graph Models (ERGMs) (Frank & Strauss, 1986; Frank, 1991; Hoff, 2005; Hunter, 2007; Snijders et al., 2006; Wasserman & Pattison, 1996).

Researches on network monitoring can be classified into two categories, i.e. structural-statistic-based and model-based methods. In the category of structural-statistic-based methods, Priebe, Conroy, Marchette, and Park (2005) proposed a scan method for sizes of k^{th} order neighorhoods. McCulloh and Carley (2011) adopted the Exponentially Weighted Moving Average (EWMA) and CUmulative SUM (CUSUM) control charts for average closeness and betweenness metrics. Neil, Hash, Brugh, Fisk, and Storlie (2013) developed a scan method for local out-stars and k-paths. Perry (2020) proposed an EWMA-based control chart for detecting the event of less mutual dyads and more transitive triplets.

In the category of model-based methods, Heard, Weston, Platanioti, and Hand (2010) modeled communication counts over time as conjugate Bayesian models and detected anomalies based on predictive p-values. Sparks (2015), Sparks (2016) and Sparks and Wilson (2019) proposed EWMA- and CUSUM-type control charts for communication counts smoothed by temporal EWMA models. Azarnoush, Paynabar, Bekki, and Runger (2016) applied a likelihood-based method to monitor the logistic regression for edge existences. Zou and Li (2017)

* Corresponding author at: College of Management and Economics, Tianjin University, No.92 Weijin Road, Nankai District, Tianjin 300072, China.

https://doi.org/10.1016/j.cie.2020.106451
Received 31 July 2019; Received in revised form 25 February 2020; Accepted 30 March 2020
Available online 06 April 2020
0360-8352/



(a) April 9, 2001 (b) May 14, 2001

(c) May 21, 2001 (d) December 17, 2001

Fig. 1. Enron email communication networks of CEOs, presidents, and vice presidents in the weeks around April 9, May 14, May 21, and December 17 in the year 2001.

proposed a Singular Value Decomposition (SVD)-based method to detect the deviation matrix based on the Network State Space Model (NSSM). Fotuhi, Amiri, and Maleki (2018) proposed the extended T^2, F and a standardized likelihoood ratio test method to monitor Poisson regression models for edge counts. Yu, Woodall, and Tsui (2018) and Wilson, Stevens, and Woodall (2019) used composite Hotelling T^2 and Shewhart control charts to detect changes in the Degree-Corrected Stochastic Block Model (DCSBM). Dong, Chen, and Wang (2019) proposed a score-test method to monitor the multilayer weighted stochastic block model.

Structural-statistic-based approaches impose no dependence assumptions and are easily computable. In previous studies, network features considered were mostly limited to the density and the degree variability. Moreover, anomalies of different features were detected separately without considering their correlations, which can be misleading. The model-based approaches characterized the network formation mechanism by probabilistic models. Communications between node pairs are assumed to depend on node attributes, fixed global or local block structure, or latent distances, which limit their applications. Besides, degeneracy and computational complexity increase the difficulty in practical use of model-based approaches (Handcock, Robins, & Snijders, 2003). In this paper, we focus on such cases that a regular communication pattern may occur within different parts of nodes over time. The significant network features including the density, reciprocity, degree variability, and transitivity are taken in account. The count statistics of the edges, mutual dyads, stars, and triangles are monitored simultaneously by a Hotelling T^2 control chart with their correlations considered. To validate the effectiveness of the proposed method, a performance evaluation framework is provided with normal and abnormal network features simulated through ERGMs. It should be noted that there is a similar study by Perry (2020), who did very excellent work on identifying the hierarchical tendency. Different from our motivation, the anomaly of his interest is the specific event of lower mutuality and higher transitivity. As will be shown in the numerical experiments, his EWMA-based control chart fails to detect the increase of mutual dyads and is insensitive to the shifts of degree variability and transitivity parameters.

The contributions of this paper are as follows. First, this paper comprehensively considers the important features of social networks and provides statistic candidate sets for both undirected and directed networks. Second, a multivariate control chart is adopted to monitor the structural statistics simultaneously so as to account for their

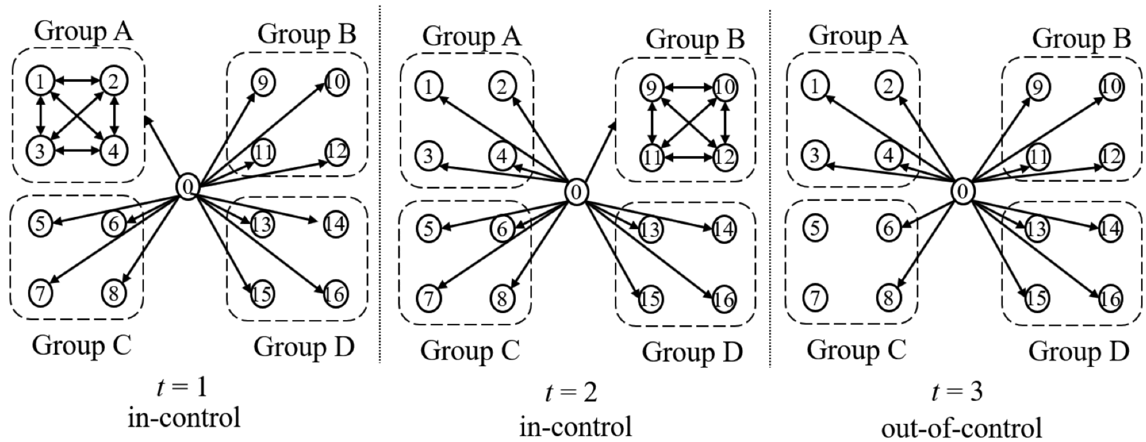


Fig. 2. An example of in-control and out-of-control status in a class network.

correlations and to decrease the overall false alarm rate. Third, we provide a performance evaluation framework based on the ERGMs in order to simulate the shifts of density, reciprocity, degree variability, and transitivity. The results of the numerical experiments show that the Hotelling T^2 control chart for the structural statistics outperforms several benchmark methods especially in detecting the large shifts of reciprocity and transitivity.

The organization of this paper is as follows. Section 2 describes the background and motivations. Section 3 presents the methods for calculating structural statistics for undirected and directed networks as well as the monitoring strategy. In Section 4, the ERGM-based performance evaluation framework is provided. In Section 5, the performance of the proposed method is compared with several benchmark methods. The Enron email network data are analyzed to illustrate the effectiveness of the proposed method in Section 6. Conclusion and future researches are given in Section 7.

2. Background and motivations

It's common in practice that there is a regular communication pattern over time while active nodes are different. To clarify this type of anomalies studied in this paper, we introduce the background and motivations through a real case and a toy example in this section. Then selection of structural statistics for characterizing the patterns is introduced.

2.1. The Enron email network

The Enron email network is a typical social network that has been widely studied. We adopt the version of data from Priebe et al. (2005), which contains the email communications for 184 unique email addresses of 150 users (mostly executives and some assistants and traders) from November 1998 to June 2002. Taking the email addresses as the nodes, directed edges are constructed if there is at least one email from the sender to the receiver in a week.

To illustrate the problem, we consider the subnetwork composed of CEOs, presidents, and vice presidents. Networks from 4 weeks in the year 2001 are presented in Fig. 1. Figs. 1(a–b) correspond to the weeks around April 6 and May 14, when the company was operating normally. The communication patterns are similar with emails sent from a few nodes to many receivers. In Fig. 1(a), communications occur among 17 nodes including the nodes 8, 12, 13, 15, 16, 19, 22, 23, 26, 27, 28, 31, 32, 34, 35, 38, 39. Emails are mostly sent from nodes 16, 23, and 38. In Fig. 1(b), communications occur among 21 nodes including nodes 2, 7, 8, 10, 12, 13, 14, 16, 18, 19, 23, 26, 27, 28, 31, 32, 33, 35, 36, 37, 38. Emails are mostly sent from nodes 32 and 37. There are 13 overlapping active nodes and 12 nodes are active in only one of the networks. In the meanwhile, the "center" nodes sending out emails in the two weeks are different. Although the active nodes are different, the communication pattern and overall communication occurrences are similar. Both networks are normal. Fig. 1(c) corresponds to the week around May 21. The network has a more outspreading communication pattern with apparently more "center" nodes and more communication occurrences compared with Figs. 1(a–b). This network is an anomaly, which is right after the "Secret" meeting at Peninsula Hotel in LA by the leaders on May 17, 2001. Fig. 1(d) corresponds to the week around December 17, which is apparently another anomaly. The outspreading communication pattern is broken. Very few emails are sent among few nodes. Such anomalous pattern occurred in 2 weeks after Enron filed for Chapter 11 bankruptcy protection.

2.2. A project representation network

To more clearly illustrate the applications, we present a toy example of project representations in a class. Consider a simplified class network with a teacher and 16 students. Index the teacher as node 0 and the students as nodes 1–16. Students are classified into 4 groups A, B, C, and D to work on different projects. In each week, only 1 group is arranged to present their project. In the week when the i^{th} group is due to present, students in the i^{th} group discuss with each other before the class. In each week, the teacher regularly lectures to all the students, showing an outreaching star structure from node 0 to nodes 1–16. Fig. 2 illustrates the communication pattern of the class in different weeks. Group A and Group B respectively represent their projects in the first 2 weeks. Mutual communications occur in Group A in the 1st week and in Group B in the 2nd week. Although the locations of the mutual communications are different, networks at $t = 1$ and $t = 2$ have the same structures and are both in-control. In the 3rd week, two students (nodes 5 and 7) in Group C are absent. The intensive mutual communication pattern are not present in the network, indicating an anomaly.

Through the Enron case and this toy example, we show that our interest is the anomaly detection of the communication patterns, which can be characterized by the occurrences of featured local structures such as the star, mutual dyads, and triangles. The changes of locations of communication patterns, i.e. different active nodes, may not induce an anomaly as long as the overall interactive mode remains unchanged. Such phenomenon is referred to as the exchangeability in the study of social network analysis modeling, meaning relabeling the nodes does not change the numbers of the structures (Lauritzen, Rinaldo, & Sadeghi, 2018).

2.3. Selection of structural statistics

Under this background, our next question is what structural statistics should be adopted for analysis. Extensive empirical and theoretical studies have shown that the density, the reciprocity, the degree variability, and the transitivity are the most common and the very important features for characterizing the communication patterns in social networks (Barabási & Albert, 1999; Bierstedt & Blau, 1965; Davis, 1970; Frank & Strauss, 1986; Holland & Leinhardt, 1981; Simmel, 1950; Snijders et al., 2006). The density is the most basic property of a network, which can be characterized by the total number of interactions. The degree variability reflects the heterogeneity of nodes, corresponding to the variance of the node degrees. It can be characterized by the number of 2-star structures when the density is fixed (Snijders, 1981). The transitivity is often interpreted as "friends of mine are my friends", reflecting the clustering effects among triplets. It can be characterized by the number of triangles or transitive triplets (Davis, 1970). The reciprocity reflects the tendency of mutual communications in a directed network and can be characterized by the number of mutual dyads (Bierstedt & Blau, 1965). The calculation of the frequently-used structural statistics will be introduced in detail in the next section.

The dominant features of a network may differ from one to another. For small-size networks like the example here, visualized graphs can be of much assistance to choose proper structural statistics. For networks with a large number of nodes, a quantitative method is necessary for choosing the most proper statistics. The presence of a certain structure may be because it is a dominant feature or simply randomly formed due to other dominant structures. For instance, the presence of a triangle formed by nodes 1, 2, and 3 may be due to the presence of the star 1–2–3 and the star 2–1–3 (the concept of the structures will be explained in detail in Section 3). Therefore, to determine whether a structural statistic plays a significant role, a hypothesis testing method is needed. One possible approach is to employ Monte Carlo simulations to test the significance of the number of the structures, which is a direction for our future research. In the current study, we focus more on the power of simultaneously monitoring the featured structural statistics compared to other existing methods. Empirically, structural statistics chosen based on the commonly-used features tend to well characterize the communication patterns in a social network.

3. Monitoring structural statistics for undirected and directed networks

In this section, we introduce the structural statistics for characterizing the density, reciprocity, degree variability, and transitivity features for both undirected and directed networks. Then the monitoring strategy is provided.

3.1. Structural statistics for undirect networks

For undirected networks, the density, degree variability and transitivity features can be basically characterized by the numbers of local structures including edges, 2-stars, and triangles. These structures are illustrated in Fig. 3. The edge structure describes whether there are interactions between two nodes. A node and its connected edges form a

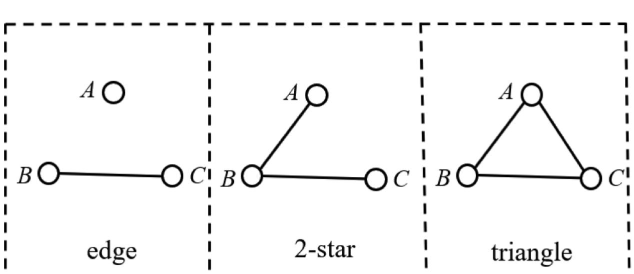


Fig. 3. The edge, 2-star and triangle structures among three nodes in an undirected network.

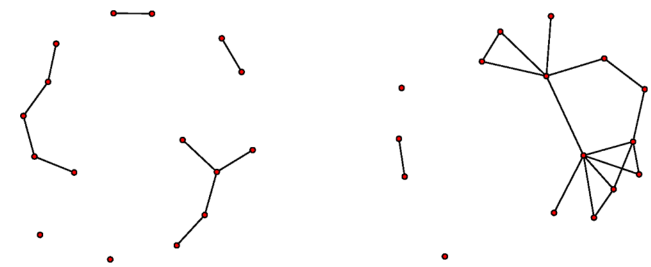


Fig. 4. A network with more stars (left) and a network with more triangles (right).

star structure. That is, a k-star contains a node and k edges connected with it. The number of k-stars can be equivalently derived from the degree distribution as $\sum_i d_i C_i^k$, where d_i is the frequency of degree i and C is the combination operator. When k takes the value 1, the 1-star structure is in fact an edge. It is possible that higher order star structures are dominant. Here we show the basic 2-stars as an example. A triangle means all three nodes are connected with each other, which describes the closure effect in triplets. Through the numbers of edges, stars, and triangles, the main features of a network are summarized (Snijders et al., 2006).

As discussed in Section 2.3, dominant features of a small-size network can be selected based on visualized graphs. Fig. 4 shows two examples of networks having dominant star structures and triangular structures respectively. For large-scale networks, we suggest the adoption of the numbers of edges, 2-stars and triangles as a start, which have proved to be widely applicable in empirical studies (Lusher, Koskinen, & Robins, 2013; Snijders et al., 2006).

Denote $S_1(Y)$, $S_2(Y)$ and $T(Y)$ as the numbers of edges, 2-stars and triangles in network Y, respectively. The count statistics are calculated as

$$\begin{aligned} S_1(Y) &= \sum_{1\leqslant i<j\leqslant n} Y_{ij} && \text{number of edges,} \\ S_2(Y) &= \sum_{1\leqslant i\leqslant n} \binom{Y_{i+}}{2} && \text{number of } 2-\text{stars,} \\ T(Y) &= \sum_{1\leqslant i<j<h\leqslant n} Y_{ij}Y_{ih}Y_{jh} && \text{number of triangles,} \end{aligned} \tag{1}$$

where i, j, h are node indexes; the $+$ sign denotes summation over the index, and Y_{i+} is the degree of node i (Frank & Strauss, 1986).

3.2. Structural statistics for direct networks

The edge direction provides valuable information of node interactions. Taking the degree measure as an example, a higher in-degree shows higher "popularity" of a person, and a higher out-degree indicates higher "friendliness" of a person. While it is sometimes true, it is not necessary that a "popular" person is friendly. Monitoring networks considering edge direction can provide us important insights of changes of the interacting pattern within the networks, which may be missed if networks are dealt with as undirected data. A "mixed-2-star" is such a structure that a node has one edge directed toward it and the other edge directed away from it. It reflects the correlations between in- and out-degrees (Lusher et al., 2013).

Types of structures within three nodes in directed networks are shown in Fig. 5. Similar to undirected networks, the total number of edges represents the overall density of a directed network. In addition, the number of reciprocated edges indicates the mutuality between nodes in the network. In Fig. 5, both the edge from A to B and the edge from B to A exist, meaning the interaction between A and B is mutual. The types of directed stars include the in-2-star, out-2-star and mixed-2-star as shown in Fig. 5. A directed triangle can be either transitive or cyclic. Given that B is a friend to A and C is a friend to B, if C is a friend

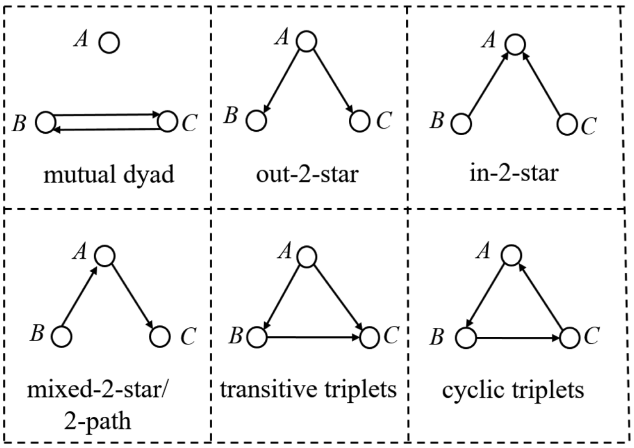


Fig. 5. Directed counterparts of the edge, 2-star and triangle structures among three nodes in a directed network.

of A, then they comprise transitive triplets; if A is a friend to C, then they comprise cyclic triplets. The numbers of edges, reciprocated edges, in-2-stars, out-2-stars, mixed-2-stars, transitive triplets and cyclic triplets reflect the density, mutuality, in-degree variability, out-degree variability, potential transitivity, transitivity and cyclicity, respectively (Robins, Pattison, & Wang, 2009; Snijders et al., 2006; Wasserman & Pattison, 1996).

Denote $S_1(Y)$, $M(Y)$, $S_2^{in}(Y)$, $S_2^{out}(Y)$, $S_2^{mix}(Y)$, $Tr(Y)$ and $Cy(Y)$ as the numbers of all edges, mutual dyads, in-2-stars, out-2-stars, mixed-2-stars, transitive triplets and cyclic triplets, respectively. Corresponding count statistics can be calculated as

$$
\begin{aligned}
S_1(Y) &= \sum_{i,j} (Y_{ij} - Y_{ij}Y_{ji}) && \text{number of edges}\\
M(Y) &= \frac{1}{2}\sum_{i,j} Y_{ij}Y_{ji} && \text{number of mutual dyads}\\
S_2^{in}(Y) &= \sum_{1\leqslant i\leqslant n} \binom{Y_{+i}}{2} && \text{number of in} - 2 - \text{stars}\\
S_2^{out}(Y) &= \sum_{1\leqslant i\leqslant n} \binom{Y_{i+}}{2} && \text{number of out} - 2 - \text{stars}\\
S_2^{mix}(Y) &= \sum_{i,j,h;i\neq h} Y_{ij}Y_{jh} && \text{number of mixed} - 2 - \text{stars}\\
Tr(Y) &= \sum_{i,j,h} Y_{ij}Y_{jh}Y_{ih} && \text{number of transitive triplets}\\
Cy(Y) &= \frac{1}{3}\sum_{i,j,h} Y_{ij}Y_{jh}Y_{hj} && \text{number of cyclic triplets,}
\end{aligned}
\tag{2}
$$

where i, j, h are node indexes; Y_{+i} and Y_{i+} are in- and out-degrees.

3.3. The monitoring strategy

The structural statistics for the random network Y can be written in a vector form as $\mathbf{g}(Y)$. To clarify the concept of the structural statistic vector, we consider a simple example as shown in Fig. 6. In the undirected network in Fig. 6(a), there are 8 edges, i.e. edges 1–2, 1–3, 1–4, 1–5, 1–6, 5–6, 7–5, and 7–6. For the 2-stars, the degrees of nodes 1 to 7 are 5, 1, 1, 1, 3, 3, and 2. Therefore, the number of 2-stars is equal to $C_5^2 + C_3^2 + C_3^2 + C_2^2 = 17$. For the triangles, we consider all possible combinations of three nodes and only those having three edges can form a triangle. In this case, the number is 2. One triangle is formed by edges 5–6, 6–7, and 7–5, and the other is formed by edges 1–5, 5–6, and 6–1. If we consider the structural statistic vector $\mathbf{g}(Y)$ composed by the numbers of edges, 2-stars, and triangles, then the value of $\mathbf{g}(Y)$ is (8, 17, 2). In the directed network in Fig. 6(b), the total number of edges is 8 as well, i.e. the directed edges 1→2, 1→3, 1→4, 1→5, 6→1, 5→6, 5→7, and 7→6. For the out-2-stars, only node 1 and node 5 have at least 2 out-edges. The number of out-2-stars for node 1 is $C_4^2 = 6$ and the numbers of out-2-stars for node 5 is 1 (7←5→6). For the mixed-2-stars, nodes 1, 5, 6, and 7 have both in- and out- edges. The number of mixed-2-stars for nodes 1, 5, 6, and 7 are 4, 2, 2, and 1. There is an in-2-star (5→6←7), a transitive triplet (5→7→6 and 5→6), and a cyclic triplet (1→5→6→1). The structural statistic vector $\mathbf{g}(Y)$ composed of the numbers of in-2-stars, mixed-2-stars, out-2-stars, transitive triplets, and cyclic triplets is (1, 9, 7, 1, 1).

In a stable network process, network samples are expected to be drawn from the same population. The monitoring of networks is essentially to test the null versus alternative hypotheses

$$
\begin{aligned}
&H_0\colon E(\mathbf{g}(Y_1)) = E(\mathbf{g}(Y_2))=...=E(\mathbf{g}(Y_T)) = \boldsymbol{\mu}_{\mathbf{g}},\\
&H_1\colon E(\mathbf{g}(Y_t)) \neq \boldsymbol{\mu}_{\mathbf{g}},
\end{aligned}
\tag{3}
$$

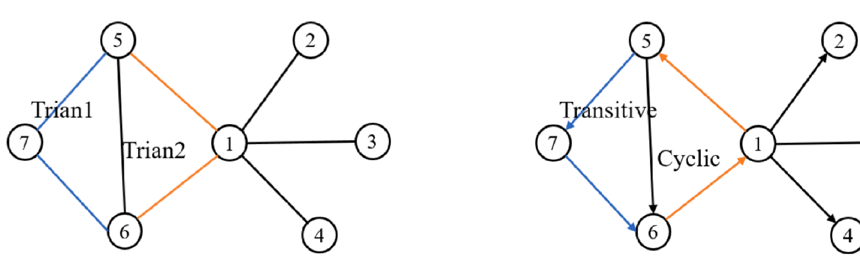
where $\boldsymbol{\mu}_{\mathbf{g}}$ is a constant equal to the expectation of the statistic vector $\mathbf{g}(Y)$ when the network process is in control. Changes in any component of $\mathbf{g}(Y)$ induce an outlier. Denote the network sample at time t as y_t ($t = 1, 2, ..., T$) and corresponding structural statistic vector as $\mathbf{g}(y_t)$. The Hotelling T^2 statistic for $\mathbf{g}(y_t)$ is given as

$$
T_t^2 = (\mathbf{g}(y_t) - \boldsymbol{\mu}_{\mathbf{g}})'\Sigma_{\mathbf{g}}^{-1}(\mathbf{g}(y_t) - \boldsymbol{\mu}_{\mathbf{g}}),
\tag{4}
$$

where $\boldsymbol{\mu}_{\mathbf{g}}$ and $\Sigma_{\mathbf{g}}$ are the expectation and variance-covariance matrix for $\mathbf{g}(Y)$. In practice, $\boldsymbol{\mu}_{\mathbf{g}}$ and $\Sigma_{\mathbf{g}}$ are usually unknown and need to be estimated. A natural and simple way is to estimate them as the mean and covariance of the samples $\mathbf{g}(y_1), ..., \mathbf{g}(y_T)$, i.e. $\hat{\boldsymbol{\mu}}_{\mathbf{g}} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{g}(y_t)$ and $\hat{\Sigma}_{\mathbf{g}} = \frac{1}{T-1}\sum_{t=1}^{T}(\mathbf{g}(y_t) - \hat{\boldsymbol{\mu}}_{\mathbf{g}})(\mathbf{g}(y_t) - \hat{\boldsymbol{\mu}}_{\mathbf{g}})'$. Notice that the structural statistics are count data. We assume the vector $\mathbf{g}(Y_t)$ with k components follows a multivariate Poisson distribution. Then, the statistic T^2 in Eq. (4) with the estimates inserted has an approximate χ^2 distribution with k degrees of freedom as studied in Patel (1973). Therefore, the control limit is given as

$$
UCL = \chi_\alpha^2(k),
\tag{5}
$$

where α is the probability of false alarm, and k is equal to the number of structural statistics.



(a) An undirected network (b) A directed network

Fig. 6. Examples for calculating structural statistics of undirected and directed networks.

4. Performance evaluation framework based on exponential random graph models

As stated in Section 2, the density, reciprocity, degree variability, and transitivity are important features describing the communication patterns. To assess whether a network control chart is capable of timely detecting the shifts in these features, we provide an ERGM-based performance evaluation framework here. Performances of the proposed method and other methods will be evaluated through the proposed framework in Section 5.

4.1. Network simulation

To evaluate the performance of a network control chart, two stages are involved. The first is Phase I estimation, including setting up an in-control model, simulating Phase I samples, and estimating process parameters and control limits. The second is Phase II monitoring, including shifting model parameters to simulate out-of-control scenarios, generating out-of-control samples from a certain time point, calculating chart statistic and comparing it with the control limits, recording the Run Lengths (*RLs*) until detecting an outlier, and calculating the Average *RLs* (*ARLs*). In either stage, simulating network samples plays an indispensable role. To make the performance evaluation framework more presentable, we first describe the exponential random graph model and MCMC-based simulation approach here.

Denote a random graph as Y and its observation as y. The space containing all possible realizations of the random network Y is denoted as $\mathcal{Y}$. The probability of y is assumed to be dependent on its structural statistic vector $\mathbf{g}(y)$. The ERGM is written as

$$P(Y=y|\theta)=\frac{exp(\theta'\mathbf{g}(y))}{k(\theta,\mathcal{Y})}, \tag{6}$$

where θ is the parameter vector corresponding to the statistic vector $\mathbf{g}(y)$, and $k(\theta,\mathcal{Y})$ is the normalizing constant equal to the summation of the numerator counterparts over all possible samples of Y, i.e. $k(\theta,\mathcal{Y})=\sum_{y^*\in\mathcal{Y}}\exp(\theta'\mathbf{g}(y^*))$ (Frank & Strauss, 1986; Frank, 1991; Hunter, 2007; Snijders et al., 2006; Wasserman & Pattison, 1996). The parameter θ can be interpreted as the tendency of the presence of local structures specified by $\mathbf{g}(y)$. Parameter estimation algorithms can be referred to in Snijders (2002), Hunter and Handcock (2006), Hunter, Krivitsky, and Schweinberger (2012).

The idea of the MCMC-based simulation approach is to construct a Markov chain $Y^{(1)}, Y^{(2)}, \ldots, Y^{(t)}, \ldots$with the stationary distribution converging to the exponential random graph distribution. Denote the transition probabilities as

$$P(y^a,y^b)=P\{Y^{(t+1)}=y^b|Y^{(t)}=y^a\}, \tag{7}$$

for $y^a, y^b\in\mathcal{Y}$. To guarantee the uniqueness of the stationary distribution π, the transition probability is required to satisfy (Snijders, 2002)

$$log(\frac{P(y^a,y^b)}{P(y^b,y^a)})=\theta'(\mathbf{g}(y^b)-\mathbf{g}(y^a)). \tag{8}$$

The key part is the updating rule in moving from the current graph $y^{(t)}$ to the next graph $y^{(t+1)}$. The updating rule consists of choosing a pair of nodes i and j at random and removing or adding an edge between them according to whether there are already an edge. If the probability of $y^{(t+1)}$ is smaller than the probability of the old one, we only sometimes accept the proposed change to the graph, with a probability that depends on the ratio of how much likely $y^{(t)}$ is than $y^{(t+1)}$. The Metropolis-Hastings (M-H) algorithm is described as follows (Hastings, 1970; Snijders, 2002).

1. *Proposal*. Given $y^{(t)}$, draw a proposal at step $t+1$ as $y\sim P(\cdot|y^t)$, where

$$P(y|y^{(t)})=\frac{(\theta'\mathbf{g}(y))}{\sum_{\tilde{y}\in\mathcal{Y}^{(t+1)}(I)}\exp(\theta'\mathbf{g}(\tilde{y}))}, \tag{9}$$

and the notation I refers to a subset of $\{(i,j)|i\neq j\}$,i.e. the set of elements of the adjacency matrix to be updated in step t; and $\mathcal{Y}^{(t+1)}(I)$ the set of adjacency matrices with elements equal to y_{ij}^t for $(i,j)\notin I^{(t+1)}$. The set $\mathcal{Y}^{(t+1)}(I)$ is the set of all allowed outcomes of $Y^{(t+1)}$, containing $2^{|I|}$ elements with $|I|$ being the size of $|I|$.
2. *Updating*. Accept the proposal $Y^{(t+1)}=y$ with probability $min(\alpha(y_t,y),1)$. Otherwise set $Y^{(t+1)}=y^{(t)}$. Here,

$$\alpha(y^{(t)},y)=\frac{\pi(y)P(y|y^t)}{\pi(y^t)P(y^t|y)}. \tag{10}$$

The network at step $t+1$ is updated as

$$y^{(t+1)}=\begin{cases}y, & \text{if } \alpha(y^{(t)},y)>u,\\ y^t, & \text{otherwise},\end{cases}. \tag{11}$$

where u is a sample drawing from the standard uniform distribution, $u\sim U[0,1]$.
3. *Obtaining samples at convergence*. Iteralize the updating step until convergence. Samples can be drawn at every other H steps, with H large enough to leave very weak correlation between successive samples.

4.2. Performance evaluation framework

The two-phase ERGM-based performance evaluation framework is presented as follows.

Phase I estimation

Step 1. Initial setup. Set the parameters for the in-control model. Since it is often not clear whether a model parameter vector is proper, we suggest starting from estimating parameters for an initial network y^0 consisting of n nodes with certain structures. Denote the ERGM parameter vector estimated for y^0 as θ and denote structural statistic vector as $\mathbf{g}^0$.

Step 2. Network generation in in-control scenario. Randomly draw T network samples through the M-H simulation algorithm.

Step 3. Estimating control limits. Calculate the structural statistics for each network sample and obtain the sequence of statistic vectors $\mathbf{g}^1, \mathbf{g}^2, \ldots, \mathbf{g}^T$. Estimate the mean and variance-covariance matrix based on the T samples as $\hat{\mu}_g=\frac{1}{T}\sum_{t=1}^T\mathbf{g}^t$ and $\hat{\Sigma}_g=\frac{1}{T-1}\sum_{t=1}^T(\mathbf{g}^t-\hat{\mu}_g)(\mathbf{g}^t-\hat{\mu}_g)'$. Calculate the control limit based on Eq. (5).

Phase II monitoring

Step 1. Initial setup. Set the change point as τ. Determine the model parameter with potential shifts after time τ and denote the shifted parameter as θ^*.

Step 2. Network generation. Draw a random network sample at time t as

$$y^t\sim\begin{cases}P(Y=y|\theta), & \text{if } t\leqslant\tau,\\ P(Y=y|\theta^*) & \text{if } t>\tau.\end{cases} \tag{12}$$

Calculate network statistic $\mathbf{g}^t$ and calculate the Hotelling T^2 statistic as T_t^2 by Eq. (4). Compare T_t^2 with the *UCL* obtained from Phase I. Stop network sampling and record $t-\tau$ as the *RL* if $T_t^2>UCL$. Otherwise, draw a new sample for time $t+1$.

Step 3. Evaluating ARLs. Repeat the network generation step for N times and obtain N *RLs*. Calculate *ARL* as $ARL=\frac{1}{N}\sum_{i=1}^N RL_i$ to evaluate the performance of the network control chart.

The above framework is applied to the evaluation of the proposed Hotelling T^2 control chart for network structural statistics in detecting shifts of ERGM parameters. It can be easily extended to the evaluation of other network control charts by replacing the structural statistic calculation and control limit estimation parts with their counterparts. This simulation framework will be adopted in the next section for

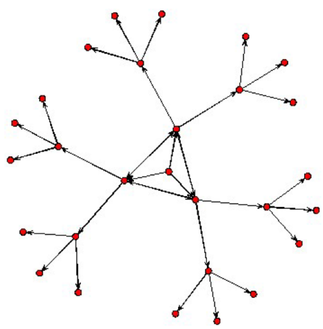


Fig. 7. The company communication network.

comparing different control charts in detecting the shifts of density, reciprocitydegree variability, and transitivity.

5. Numerical experiments for performance comparison

To validate the effectiveness of the proposed monitoring strategy, we compare our method with the EWMA-based control chart proposed by Perry (2020), which is designed for detecting the hierarchical tendency present in directed networks. The EWMA control charts for the average and standard deviation of the degree measure in Hosseini and Noorossana (2018) are also considered as benchmarks with a slight modification of the degree measure into out-degrees. Shifts of mutuality, out-degree variability and transitivity are simulated as the out-of-control scenarios through the proposed ERGM-based performance evaluation framework.

We consider a communication network in a company with information distributed from directors to employees, which presents an outspreading pattern as shown in Fig. 7. Mutual and transitive communications occur among the directors. Here, the reciprocity, out-degree variability and transitivity features are considered. An ERGM with the numbers of mutual dyads, out-2-stars, and transitive triplets as the modeling terms is fitted to this directed network as the in-control model. The model parameters are 0.25, -1.68, and 0.69. Control limits are obtained to achieve an ARL_0 of 370 based on 1000 simulation runs for the four control charts. As for the EWMA control charts, the commonly-used values of 0.05, 0.10, and 0.20 are chosen for the smoothing parameter λ. For the out-of-control scenarios, we consider the increases of the reciprocity, out-degree variability, and transitivity. The model parameter for the number of mutual dyads is increased by 0.20, 0.40, and 0.60; the parameter for the number of out-2-stars is increased by 0.10, 0.20, and 0.30; and the parameter for the number of transitive triplets is increased by 0.05, 0.10, and 0.15. Run lengths with values larger than 1000 are returned as 1000. The *ARL*s corresponding to the different control charts are listed in Table 1. The smallest *ARL* for each out-of-control scenario is bolded. The ARL_0 values are reported in the "none" scenario, which indicates no shifts in the network features.

From Table 1, we can see that the Hotelling T^2 control chart for the numbers of mutual dyads, out-2-stars, and transitive triplets performs comparably well with the EWMA control chart for average degree with λ of 0.05 in most scenarios. EWMA control chart for average degree is more efficient in detecting the shifts of out-degree variability and small shifts of reciprocity and transitivity. By contrast, the Hotelling T^2 control chart for the structural statistics detects the large shifts of reciprocity and transitivity more promptly. As for the EWMA-based control chart for hierarchical tendency, the underlined *ARL*s indicate the performance largely decreases with the increase of the mutual dyads. This is because Perry's control chart is designed to detect the hierarchical tendency, which is the occurrence of both reduced mutual dyads and increased transitive triplets. In these scenarios, the number of mutual dyads increases, leading to a lower alarm probability.

In this section, we have investigated the performances of several monitoring methods for directed networks. It is noticeable that the proposed ERGM-based performance evaluation framework can also be extended for evaluating methods for monitoring undirected networks. Although not presented here, we have conducted extensive numerical experiments for undirected networks. Methods were compared including the Hotelling T^2 control chart for the number of edges, 2-stars, and triangles, the DCSBM-based approach by Wilson et al. (2019), and the EWMA control charts for the average and the standard deviation of degree measure by Hosseini and Noorossana (2018). The Hotelling T^2 control chart for the structural statistics is shown to perform the best in detecting the shifts of transitivity and perform satisfactorily in detecting the medium and large shifts of density and degree variability.

6. Application to the Enron email communication network

We illustrate the proposed method by applying it to the analysis of the Enron email communication networks. We notice that a very large proportion of nodes are inactive during the early and late period, providing little meaningful information on company operation. Considering the major events including the financial scandal were largely concentrated in the year 2001, we focus on the 53 networks from that year.

Based on the plots of the Enron email networks, we found the communications show a strong outspreading pattern, with some mutual and transitive structures. Fig. 8 shows two networks from the 13^{th} and 40^{th} weeks as an example. The numbers of mutual dyads, out-2-stars, and transitive triplets are calculated to characterize the main features of the networks. We use the networks from the first 20 weeks as the Phase I data, based on which the mean and variance-covariance matrix are estimated. The Hotelling T^2 statistics are then calculated by Eq. (4). The Poisson assumption for the structural statistics and the Chi-squared

Table 1
*ARL*s corresponding to the Hotelling T^2 control chart for the numbers of mutual dyads, out-2-stars, and transitive triplets (multi-stat T^2), the EWMA-based control chart for the hierarchical tendency (Perry), the EWMA control chart for the average out-degree (average degree), and the EWMA control chart for the standard deviation of out-degrees (sd degree)

	multi-stat	Perry			average degree			sd degree		
	T^2	$\lambda = 0.05$	$\lambda = 0.10$	$\lambda = 0.20$	$\lambda = 0.05$	$\lambda = 0.10$	$\lambda = 0.20$	$\lambda = 0.05$	$\lambda = 0.10$	$\lambda = 0.20$
none	370.37	364.64	360.35	377.13	371.45	371.10	369.29	369.58	371.29	369.94
mutual + 0.20	200.00	*768.90*	*710.93*	*597.30*	**179.14**	228.02	266.92	328.64	378.73	346.07
mutual + 0.40	**50.00**	*966.67*	*970.83*	*857.67*	80.92	116.83	134.44	326.02	362.18	332.15
mutual + 0.60	**29.41**	*966.67*	*985.00*	*981.93*	37.97	52.08	71.92	292.79	313.88	345.07
out-2-star + 0.10	79.37	177.93	156.73	141.43	**11.53**	13.76	14.23	67.99	87.52	105.18
out-2-star + 0.20	16.95	123.27	102.13	108.87	**3.42**	3.69	4.07	21.79	24.27	27.33
out-2-star + 0.30	3.87	166.93	188.17	198.93	**1.86**	1.93	2.05	8.57	12.23	11.73
transitive + 0.05	147.06	153.70	226.10	239.33	**129.60**	169.94	186.73	298.95	419.15	330.92
transitive + 0.10	52.63	111.87	113.10	182.73	**40.21**	44.08	65.51	203.57	238.74	255.69
transitive + 0.15	**1.56**	76.10	128.17	154.57	15.92	18.69	20.52	81.75	106.97	132.87

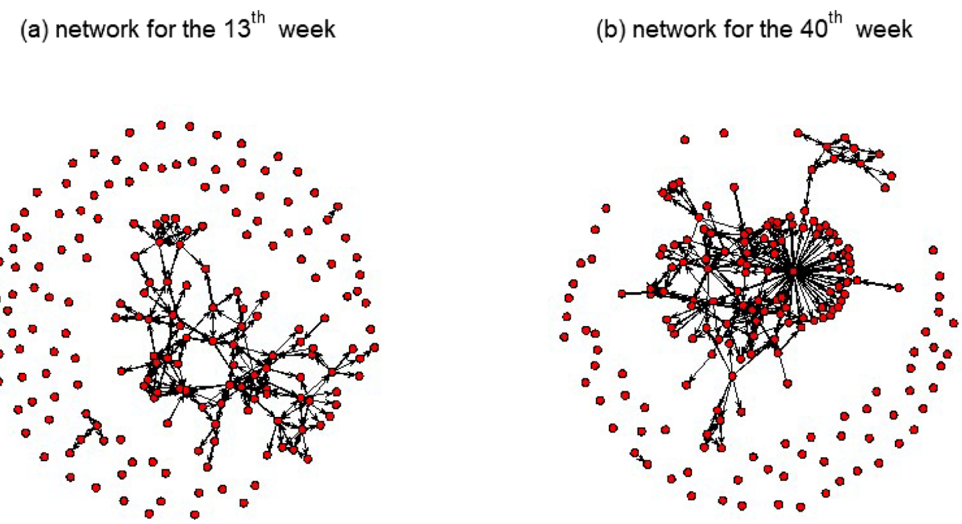
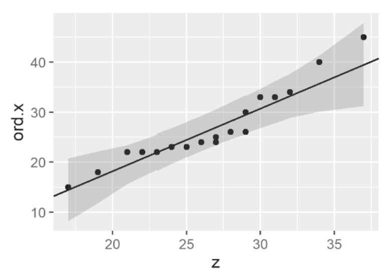
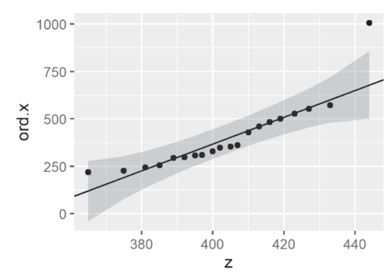


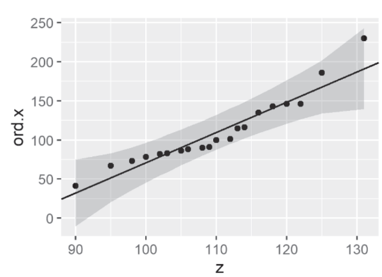
Fig. 8. The Enron email communication networks for the 13th and 40th weeks in the year 2001.



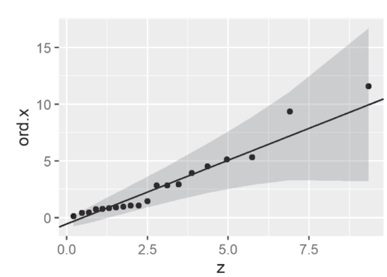
(a) Poisson distribution for the number of mutual dyads



(b) Poisson distribution for the number of out-2-stars



(c) Poisson distribution for the number of transitive triplets



(d) Chi-squared distribution for the Hotelling T^2 statistics

Fig. 9. Quantile-Quantile plots for the structural statistics and the Hotelling T^2 statistics.

distribution for the Hotelling T^2 statistic are validated by the Quantile-Quantile plots in Fig. 9. Sample points of the structural statistics generally lie within the 95% confidence interval in Fig. 9(a–c), and all points of the Hotelling T^2 statistic lie within the 95% confidence interval in Fig. 9(d). The correlations among the three statistics are significant by the Pearson correlation test with p-values smaller than 0.05. The control limit is obtained as $UCL = \chi^2_{0.9973}(3)$, i.e. 14.16.

To validate its efficiency, we compare the proposed method with the DCSBM-based method by Wilson et al. (2019) and the EWMA-based control chart for the hierarchical tendency by Perry (2020). The DCSBM-based method is applicable for weighted undirected networks. Therefore, we use the number of emails between two people as the edge weights. There is no significant pattern of communities. Hence all nodes are assumed to belong to one community. DCSBM P and δ control charts are constructed to monitor the propensity of node interaction and the variability of the interaction rate. The control charts are shown in Fig. 10, where the four subfigures (a–d) are the Hotelling T^2 control chart, Perry's EWMA-based control chart, and the Shewhart control charts for the P and δ parameters.

Many anomalies are detected through the Hotelling T^2 control chart and the Shewhart control chart for the P parameter of DCSBM, with 12 and 14 outliers respectively. The dates for the outliers are listed in Table 2. The overlapping dates for outliers detected by the T^2 and P control charts are May 21, Oct. 22, Nov. 19, Dec. 10, signalling the "secret" meeting at Peninsula Hotel in LA on May 17, the informal probe of Enron by the SEC on Oct. 17, Enron restating its third quarter

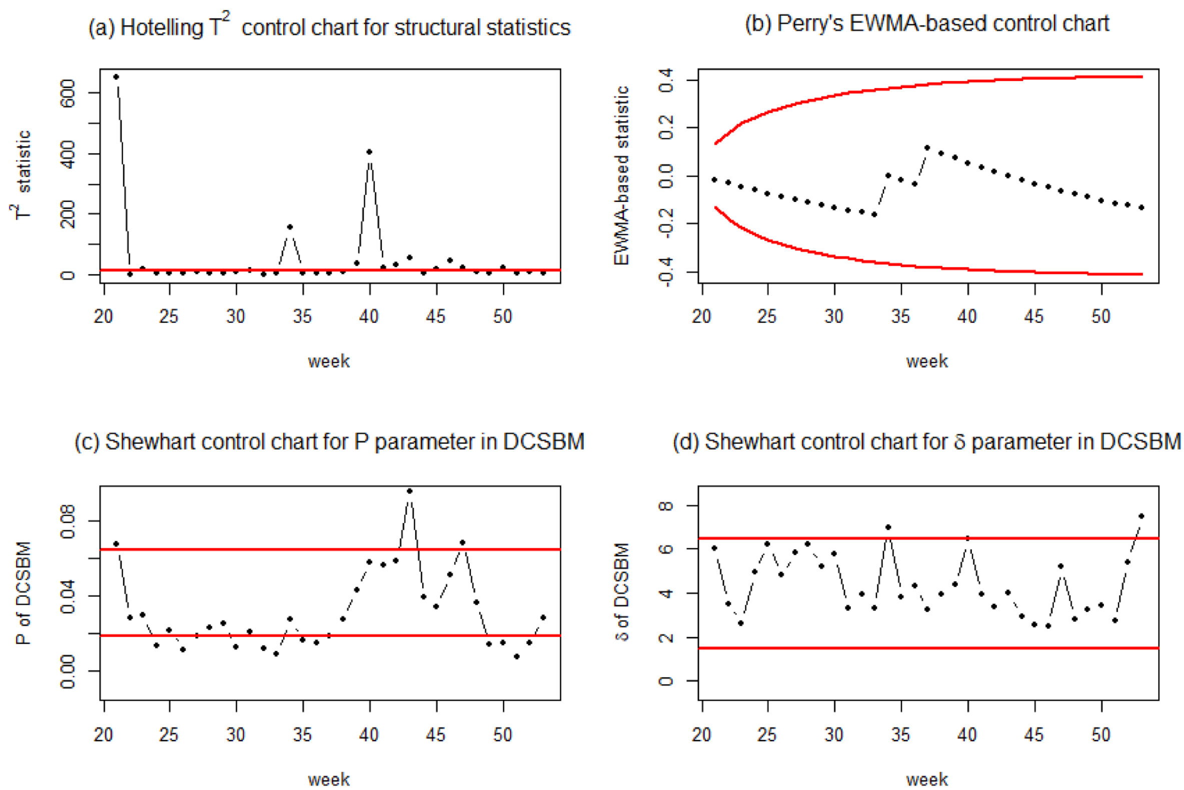


Fig. 10. The control charts for the Enron email communication network.

Table 2
Dates for the outliers detected by the Hotelling T^2 control chart for structural statistics, the EWMA-based control chart for hierarchical tendency, the Shewhart control charts for the P and δ parameters of the DCSBM

Method	Date						
T^2	5/21	6/4	8/20	9/24	10/1	10/8	10/15
	10/22	11/5	11/12	11/19	12/10		
EWMA	None						
DCSBM-P	5/21	6/11	6/25	7/23	8/6	8/13	8/27
	9/3	10/22	11/19	12/3	12/10	12/17	12/24
DCSBM-δ	8/20	12/31					

earnings on Nov. 19, and Enron filing for bankruptcy on Dec. 2. It can be seen from Table 2 that anomalies detected by the Shewhart control chart for P parameter are mostly around July and December while most outliers detected by the Hotelling T^2 control chart are around October and November. It is well known that the burst of the financial scandal is on Oct. 16 when Enron reports a $618 million third-quarter loss, followed by sequential events such as formal investigations and bankruptcy protection application. In this specific case, the Hotelling T^2 control chart for the numbers of mutual dyads, out-2-stars, and transitive triplets is sensitive enough to detect crucial events and not so sensitive as to signal insignificant issues. By contrast, the DCSBM P control chart detects some other relatively insignificant events while missed several important events. The DCSBM δ control chart and the EWMA-based control chart for the hierarchical tendency are much less effective. However, it should be noticed that Perry (2020) also studied the Enron email networks and his control chart well detected a series of important anomalies in the year 2001. The reason that his method fails in this case is that the data processing methods are different. In his study, the Enron email communication networks are aggregated on a daily basis and nodes are chosen at a wider range. In our study, the networks are aggregated on a weekly basis for a smaller group involving executives, assistants, and traders. Such different conclusions indicate the aggregation methods may contribute to an improved or reduced performance. This is a future direction worth deep exploration.

7. Conclusion and future researches

Motivated by cases where the overall communication pattern is significant while changes for specific nodes are negligible, we studied the method for detecting anomalous patterns featured by the density, reciprocity, degree variability, and transitivity. We suggested monitoring the features through the structural statistics including the numbers of edges, mutual dyads, 2-stars, and triangles. The Hotelling T^2 control chart was adopted to monitor those count statistics with their correlations considered. Our method is flexibly applicable to the monitoring of both undirected and directed networks simply by selecting the count statistics of the undirected or directed versions. An ERGM-based evaluation framework was proposed for assessing the performance of network control charts in detecting the changes of these important features.

We compared our method with benchmark methods by conducting numerical experiments following the ERGM-based evaluation framework. Control charts for comparison were the Hotelling T^2 control chart for the numbers of mutual dyads, out-2-stars, and transitive triplets, the EWMA-based control chart for the hierarchical tendency, the EWMA control chart for the average out-degree, and the EWMA control chart for the standard deviation of out-degrees. Overall, the Hotelling T^2 control chart and the EWMA control chart for average degree with λ of 0.05 perform comparably well in most scenarios. The Hotelling T^2 control chart detects the large shifts of reciprocity and transitivity more promptly. The EWMA control chart for average degree is more efficient in detecting the shifts of out-degree variability and small shifts of reciprocity and transitivity.

We applied the proposed method to the analysis of the Enron email communication networks. The data was aggregated for the year 2001 on a weekly basis. As a comparison, we applied the Shewhart control chart for the P and δ parameters of the DCSBM, for which the data were aggregated as weighted undirected networks. The EWMA-based control chart for the hierarchical tendency was also plotted. It turned out the proposed method effectively detected the important events when and after the crucial Enron financial scandal. By contrast, the Shewhart control chart for the *P* parameter signaled several relatively insignificant dates and missed some important events. The Shewhart control chart for the δ parameter and the EWMA-based control chart for the hierarchical tendency were incapable of detecting crucial events.

The main contributions of this paper are summarized as follows. First, we proposed a Hotelling T^2 control chart for multiple structural statistics that is flexibly applicable to the monitoring of undirected and directed networks. Second, we provided an ERGM-based performance evaluation framework for simulating shifts in density, reciprocity, degree variability, and transitivity. We illustrated the superiority of the proposed method through both numerical experiments and a real case. The results showed that the proposed method has a generally good performance and is exceptionally advantageous in detecting large shifts of reciprocity and transitivity.

As pointed out in previous sections, a quantitative approach to selecting dominant structural statistics and the effects of aggregation to monitoring performance are two important directions worthy of further investigation. Besides, the networks considered in this paper are binary. Extensions can be made to weighted networks. The method proposed in this paper belongs to the category of statistic-based approaches. Another possibility is to model the network samples as ERGMs and then monitor the model parameters, which belongs to the category of model-based approaches. The structural statistics adopted in this paper can be used as model terms. Although these structural statistics are sufficient statistics for ERGMs, the model-based approach may not perform as well as the proposed structural-statistic approach in detecting the anomalous communication patterns if the model is fitted to each individual network. The reason is that the structural statistics at different time points may differ a lot while the parameters may not differ in the same degree. For instance, considering an undirected network dependent on the number of triangles, the triangle counts at time $t = 8$ and time $t = 15$ are 30 and 80 respectively. The ERGM parameters for the two networks reflect the tendencies of the first network having 30 triangles and the second network having 80 triangles. Through monitoring the triangle count, the network at $t = 15$ can be easily detected. However, the difference of ERGM parameters may not be significant unless both parameters reflect the tendency of the networks having 30 triangles. To overcome the issue discussed here, a potential method is to make use of the information of all in-control sample data to estimate the ERGMs instead of fitting ERGMs individually. Corresponding monitoring strategies can be explored. In addition, anomaly detection methods for autocorrelated networks can be developed based on temporal version of the ERGMs as a future direction.

CRediT authorship contribution statement

Panpan Zhou: Methodology, Software, Writing - original draft. **Dennis K.J. Lin:** Conceptualization, Supervision, Writing - review & editing. **Xiaoyue Niu:** Methodology, Software. **Zhen He:** Supervision, Writing - review & editing, Funding acquisition.

Acknowledgments

This work was partially supported by the National Key R&D Program of China [Grant No. 2016QY02D0301] and the National Natural Science Foundation of China [Grant No. 71532008, 71672122 and 71802145].

References

Azarnoush, B., Paynabar, K., Bekki, J., & Runger, G. (2016). Monitoring temporal homogeneity in attributed network streams. *Journal of Quality Technology, 48*(1), 28–43.
Barabási, A. L., & Albert, R. (1999). Emergence of scaling in random networks. *Science, 286*(5439), 509–512.
Bierstedt, R., & Blau, P. M. (1965). Exchange and power in social life. *American Sociological Review, 30*(5), 789.
Davis, J. A. (1970). Clustering and hierarchy in interpersonal relations: Testing two graph theoretical models on 742 sociomatrices. *American Sociological Review, 35*(5), 843.
Dong, H., Chen, N., & Wang, K. (2019). Modeling and change detection for count-weighted multilayer networks. *Technometrics*. https://doi.org/10.1080/00401706.2019.1625812.
Fotuhi, H., Amiri, A., & Maleki, M. R. (2018). Phase I monitoring of social networks based on Poisson regression profiles. *Quality and Reliability Engineering International, 34*, 572–588.
Frank, O. (1991). Statistical analysis of change in networks. *Statistica Neerlandica, 45*, 283–293.
Frank, O., & Strauss, D. (1986). Markov graphs. *Journal of the American Statistical Association, 81*(395), 832–842.
Handcock, M. S., Robins, G., & Snijders, T. (2003). *Assessing degeneracy in statistical models of social networks, CSSS Working Paper no. 39.* Seattle: University of Washington.
Hastings, W. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika, 57*, 97–109.
Heard, N. A., Weston, D. J., Platanioti, K., Hand, D. J., et al. (2010). Bayesian anomaly detection methods for social networks. *The Annals of Applied Statistics, 4*(2), 645–662.
Hoff, P. D. (2005). Bilinear mixed-effects models for dyadic data. *Journal of the American Statistical Association, 100*, 286–295.
Holland, P. W., & Leinhardt, S. (1981). An exponential family of probability distributions for directed graphs. *Journal of the American Statistical Association, 76*(373), 33–50.
Hosseini, S. S., & Noorossana, R. (2018). Performance evaluation of EWMA and CUSUM control charts to detect anomalies in social networks using average and standard deviation of degree measures. *Quality and Reliability Engineering International, 34*(4), 477–500.
Hunter, D. R. (2007). Curved exponential family models for social networks. *Social Networks, 29*(2), 216–230.
Hunter, D. R., & Handcock, M. S. (2006). Inference in curved exponential family models for networks. *Journal of Computational & Graphical Statistics, 15*(3), 565–583.
Hunter, D. R., Krivitsky, P. N., & Schweinberger, M. (2012). Computational statistical methods for social network models. *Journal of Computational and Graphical Statistics, 21*(4), 856–882.
Lauritzen, S., Rinaldo, A., & Sadeghi, K. (2018). Random networks, graphical models and exchangeability. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 80*(3), 481–508.
Lusher, D., Koskinen, J., & Robins, G. (2013). *Exponential random graph models for social networks – Theory, methods, and applications.* New York, NY: Cambridge University Press.
McCulloh, I., & Carley, K. M. (2011). Detecting change in longitudinal social networks. *Journal of Social Structure, 12*(3), 1–37.
Neil, J., Hash, C., Brugh, A., Fisk, M., & Storlie, C. B. (2013). Scan statistics for the online detection of locally anomalous subgraphs. *Technometrics, 55*(4), 403–414.
Patel, H. I. (1973). Quality control methods for multivariate binomial and Poisson distributions. *Technometrics, 15*(1), 103–112.
Perry, M. B. (2020). An EWMA control chart for categorical processes with applications to social network monitoring. *Journal of Quality Technology, 52*(2), 182–197.
Priebe, C. E., Conroy, J. M., Marchette, D. J., & Park, Y. (2005). Scan statistics on enron graphs. *Computational & Mathematical Organization Theory, 11*(3), 229–247.
Robins, G. L., Pattison, P. E., & Wang, P. (2009). Closure, connectivity and degrees: New specifications for exponential random graph (p∗) models for directed social networks. *Social Networks, 31*(2), 105–117.
Savage, D., Zhang, X., Yu, X., Chou, P., & Wang, Q. (2014). Anomaly detection in online social networks. *Social Networks, 39*, 62–70.
Simmel, G. (1950). *The Sociology of Georg Simmel.* Glencoe, IL: Free Press.
Snijders, T. A. B. (1981). The degree variance: An index of graph heterogeneity. *Social Networks, 3*(3), 163–174.
Snijders, T. A. (2002). Markov chain Monte Carlo estimation of exponential random graph models. *Journal of Social Structure, 3*(2), 1–40.
Snijders, T. A. B., Pattison, P. E., Robins, G. L., & Handcock, M. S. (2006). New specifications for exponential random graph models. *Sociological Methodology, 36*(1), 99–153.
Sparks, R. (2015). *Social network monitoring: Aiming to identify periods of unusually increased communications between parties of interest.* S. Knoth, & W. Schmid *Frontiers in Statistical Quality Control 11*Cham: Springer3–13.
Sparks, R. (2016). Detecting periods of significant increased communication levels for subgroups of targeted individuals. *Quality and Reliability Engineering International, 32*(5), 1871–1888.
Sparks, R., & Wilson, J. D. (2019). Monitoring communication outbreaks among an unknown team of actors in dynamic networks. *Journal of Quality Technology, 51*(4), 353–374.
Wasserman, S., & Pattison, P. (1996). Logit models and logistic regressions for social networks: I. An introduction to markov graphs andp. *Psychometrika, 61*(3), 401–425.
Wilson, J. D., Stevens, N. T., & Woodall, W. H. (2019). Modeling and detecting change in temporal networks via the degree corrected stochastic block model. *Quality and Reliability Engineering International, 35*(5), 1363–1378.
Woodall, W. H., Zhao, M. J., Paynabar, K., Sparks, R., & Wilson, J. D. (2017). An overview and perspective on social network monitoring. *IISE Transactions, 49*(3), 354–365.
Yu, L., Woodall, W. H., & Tsui, K. L. (2018). Detecting node propensity changes in the dynamic degree corrected stochastic block model. *Social Networks, 54*, 209–227.
Zou, N., & Li, J. (2017). Modeling and change detection of dynamic network data by a network state space model. *IISE Transactions, 49*(1), 45–57.