Anomaly Detection in Time-Evolving Attributed Networks

Luguo Xue[1], Minnan Luo[1], Zhen Peng[1], Jundong Li[2],
Yan Chen[1(✉)], and Jun Liu[1,3]

[1] School of Electronic and Information Engineering, Xi'an Jiaotong University, China
luguoxuecx@gmail.com, zhenpeng27@outlook.com
{minnluo,yanchen,liukeen}@edu.xjtu.cn
[2] Computer Science and Engineering, Arizona State University, USA
jundongl@asu.edu
[3] National Engineering Lab for Big Data Analytics, Xi'an Jiaotong University, China

Abstract. Recently, there is a surge of research interests in finding anomalous nodes upon attributed networks. However, a vast majority of existing methods fail to capture the evolution of the networks properly, as they regard them as static. Meanwhile, they treat all the attributes and the instances equally, ignoring the existence of noisy. To tackle these problems, we propose a novel dynamic anomaly detection framework based on residual analysis, namely AMAD. It leverages the small smooth disturbance between time stamps to characterize the evolution of networks for incrementally update. Experiments conducted on several datasets show the superiority of AMAD in detecting anomalies.

1 Introduction

Recently, there is a surge of research focusing on anomaly detection on attributed networks, and the task is to identify the anomalous nodes whose patterns deviate from the other majority nodes in the network [4]. Particularly with the increasing use of advanced sensors and social media platforms, an increasingly amount of time-evolving data regarding attributed networks can be collected in real time. It provides us an additional dimension (*i.e.* temporal information) to analyze the evolving patterns of anomalies in attributed networks.

To this end, we study the novel problem of anomaly detection in attributed networks within a dynamic environment. Nevertheless, the problem is nontrivial to solve due to the following three challenges. First of all, as anomalous patterns may evolve in a dynamic environment, it is necessary to continuously update the previously built model in an online fashion. Secondly, since a small disturbance of network might cause a ripple effect to the derived patterns, methods need to characterize the underlying evolution mechanisms of the networks. Third, the structurally irrelevant attributes can impede us to accurately spot anomalies. Hence, identifying and filtering out these attributes is necessary.

In this paper, we propose a novel dynamic framework for anomaly detection based on residual analysis, namely AMAD. Under the assumption of temporal smoothness property [1], AMAD leverages the small evolutionary disturbance to characterize the evolution patterns of networks, and therefore update the previously results incrementally. Meanwhile, the incorporation of feature selection ensures the robustness of AMAD against the noisy features in the data. The main contributions of this work are summarized as: (1) Exploring a principled way to characterize the evolutionary patterns of networks to spot anomalies in an online fashion; (2) Formally propose a novel dynamic anomaly detection framework AMAD based on residual analysis and attribute selection; (3) Conducting experiments on several datasets and the results show the superiority of our method.

2 The Proposed Framework - AMAD

We first define the problem of anomaly detection on time-evolving neteworks and then elaborate the developed anomaly detection framework AMAD. All the notations are summarized in Table 1.

Definition: ***Anomaly detection on time-evolving attributed networks.*** *Give a time-evolving attributed network* $\mathcal{G}(t) = \{\mathcal{V}_t, \mathcal{A}_t, \mathbf{X}_t\}$ *over a series of time stamps* $t, t+1, t+2, \cdots, t+m$ $(m = 0, 1, 2, \cdots)$*, the task of anomaly detection on time-evolving attributed networks is to find a set of nodes at each time stamp that are rare and differ significantly from the majority reference nodes in the attributed network.*

Table 1: Notation Definition.

Notation	Definition	Notation	Definition
$\mathbf{A}_t \in \mathbb{R}^{n\times n}$	adjacency matrix	$\mathbf{W}_t \in \mathbb{R}^{n\times d}$	wight matrix
$\mathbf{X}_t \in \mathbb{R}^{d\times n}$	attribute matrix	$\mathbf{R}_t \in \mathbb{R}^{d\times n}$	residual matrix
$\mathcal{V}_t$	node set	$\mathbf{L}_t \in \mathbb{R}^{n\times n}$	laplacian matrix of $\mathbf{A}_t$
$\mathcal{G}(t)$	attributed network	$\alpha, \beta, \gamma, \varphi$	trade-off parameters

Modeling formulation: From the residual analysis perspective, anomalies often have a large residual value [4] and cannot well be reconstructed from the other instances in the data. According to the problem formulation of the residual analysis based anomaly detection [5], the objective function at time stamp t can be formulated as:

$$\min_{\mathbf{W}_t, \mathbf{R}_t} \mathcal{L}(\mathbf{W}_t, \mathbf{R}_t; \mathbf{X}_t) + \Omega(\mathbf{W}_t, \alpha, \beta) + \Psi(\mathbf{R}_t, \gamma, \varphi), \tag{1}$$

where the loss $\mathcal{L}(\mathbf{W}_t, \mathbf{R}_t; \mathbf{X}_t) = \|\mathbf{X}_t - \mathbf{X}_t\mathbf{W}_t\mathbf{X}_t - \mathbf{R}_t\|_F^2$. The fist regularization term on $\mathbf{W}_t$ is $\Omega(\mathbf{W}_t, \alpha, \beta) = \alpha\|\mathbf{W}_t\|_{2,1} + \beta\|\mathbf{W}_t^\top\|_{2,1}$, which is used to control the sparsity of relevant nodes and attributes. The second regularization term on $\mathbf{R}_t$ is $\Psi(\mathbf{R}_t, \gamma, \varphi) = \gamma\|\mathbf{R}_t^\top\|_{2,1} + \varphi tr(\mathbf{R}_t\mathbf{L}_t\mathbf{R}_t^\top)$, where the first term controls the sparsity of anomalies, while the second term follows the Homophily assumption that two connected nodes will be similar.

To fit the dynamic setting, we follow the temporal smoothness property and assume the optimal variables of optimization problem (1) between t and $t+1$

satisfying: $\mathbf{R}_{t+1} = \mathbf{R}_t + \Delta\mathbf{R}, \mathbf{W}_{t+1} = \mathbf{W}_t + \Delta\mathbf{W}, \mathbf{X}_{t+1} = \mathbf{X}_t + \Delta\mathbf{X}$ and $\mathbf{A}_{t+1} = \mathbf{A}_t + \Delta\mathbf{A}$, where Δ denote the small changes variables. As a result, the objective function at $t+1$ is:

$$\begin{aligned}\min_{\mathbf{W}_t+\Delta\mathbf{W},\mathbf{R}_t+\Delta\mathbf{R}} &\mathcal{L}(\mathbf{W}_t + \Delta\mathbf{W}, \mathbf{R}_t + \Delta\mathbf{R}; \mathbf{X}_t + \Delta\mathbf{X}) \\ &+ \Omega(\mathbf{W}_t + \Delta\mathbf{W}, \alpha, \beta) + \Psi(\mathbf{R}_t + \Delta\mathbf{R}, \gamma, \varphi).\end{aligned} \tag{2}$$

When the optimal $\mathbf{W}_t$ and $\mathbf{R}_t$ have been learned at t, we only need to consider the terms which containing $\Delta\mathbf{W}$ and $\Delta\mathbf{R}$. Finally, according to the triangle inequality of norms, we have:

$$\begin{aligned}\min_{\Delta\mathbf{W},\Delta\mathbf{R}} &\|\Delta\mathbf{X} - (\mathbf{X}_{t+1}\mathbf{W}_{t+1}\mathbf{X}_{t+1} - \mathbf{X}_t\mathbf{W}_t\mathbf{X}_t) - \Delta\mathbf{R}\|_F^2 + \alpha\|\Delta\mathbf{W}_t\|_{2,1} \\ &+ \beta\|\Delta\mathbf{W}_t^\top\|_{2,1} + \gamma\|\Delta\mathbf{R}^\top\|_{2,1} + \varphi tr\left((\mathbf{R}_t + \Delta\mathbf{R})\mathbf{L}_{t+1}(\mathbf{R}_t^\top + \Delta\mathbf{R}^\top)\right).\end{aligned} \tag{3}$$

To solve the problem, we employ an alternating optimization algorithm to recursively update the optimal variables. Through fixing one variable and updating another, the optimal variables $\Delta\mathbf{R}$, $\Delta\mathbf{W}$ can be solved by following equations:

$$\begin{aligned}\Delta\mathbf{R} =&(\Delta\mathbf{X} - \varphi\mathbf{R}_t\Delta\mathbf{L} - \varphi\mathbf{R}_t\mathbf{L}_t - \mathbf{X}_{t+1}\Delta\mathbf{W}_t\mathbf{X}_{t+1}) * (\mathbf{I} + \gamma\mathbf{D}_R + \varphi\mathbf{L}_t + \varphi\Delta\mathbf{L})^{-1}, \\ &\alpha\mathbf{D}_{W1}\Delta\mathbf{W} + \beta\Delta\mathbf{W}\mathbf{D}_{W2} + \sum_{\mathbf{M}\in\mathcal{M}}\sum_{\mathbf{N}\in\mathcal{N}}\mathbf{M}\Delta\mathbf{W}\mathbf{N} = \mathbf{H},\end{aligned} \tag{4}$$

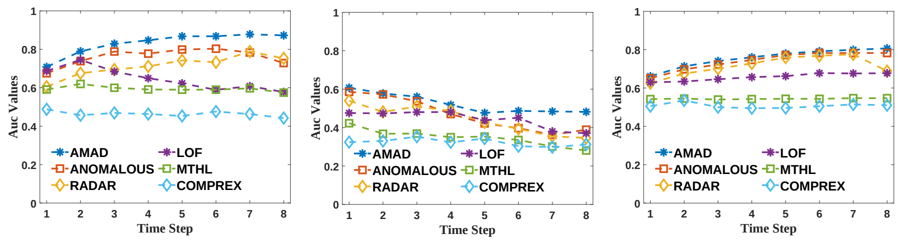
where $\mathbf{D}_R(k,k) = \frac{1}{2\|\Delta\mathbf{R}^\top(k,:)\|_2}, (k = 1, 2, \cdots, n)$, $\mathbf{D}_{W1}(k,k) = \frac{1}{2\|\Delta\mathbf{W}(k,:)\|_2}$ and $\mathbf{D}_{W2}(k,k) = \frac{1}{2\|\Delta\mathbf{W}^\top(k,:)\|_2}$ $(k = 1, 2, \cdots, n)$. The sets $\mathcal{M} = \{\mathbf{X}^\top\mathbf{X}, \mathbf{X}^\top\Delta\mathbf{X}, \Delta\mathbf{X}^\top\mathbf{X}, \Delta\mathbf{X}^\top\Delta\mathbf{X}\}$ and $\mathcal{N} = \{\mathbf{X}\mathbf{X}^\top, \mathbf{X}\Delta\mathbf{X}^\top, \Delta\mathbf{X}\mathbf{X}^\top, \Delta\mathbf{X}\Delta\mathbf{X}^\top\}$. And $\mathbf{H}$ represents the terms which not contain $\Delta\mathbf{W}$. We employ gradient descent method to solve the second equation of the problem (4).

3 Experiments

We compare AMAD with five anomaly detection methods. LOF [3], Radar [4], and ANOMALOUS [5] are static methods, while MTHL [6] and COMPREX [2] are dynamic methods. The information of datasets are listed in Table 2. And we generate time-evolving networks with anomalies by perturb their nodes.

Performance Evaluation: The anomaly detection performance is shown in Figure 1 and we adopt AUC value as metrics. We have the following observations from the figure: (1) our method achieves the best performance in majority of the time stamps, as we characterize the evolution patterns of networks and find the most relevant attributes; (2) ANOMALOUS and Radar are slightly inferior to our method as they ignore the evolutionary information of the underlying network for anomaly detection; (3) MTHL and COMPREX obtain the worst results, though they fit to dynamic setting. It emphasis the importance of instance and attribute selection.

Additionally, we compare AMAD with ANOMALOUS to demonstrate its efficiency. As shown in Table 3, AMAD can be converged faster than ANOMALOUS. It can be ascribed to that AMAD could greatly reduce the amount of computation by leveraging the sparse evolution matrices $\Delta\mathbf{X}$ and $\Delta\mathbf{A}$.



(a) Wiki (b) Blogcatalog (c) Flickr

Fig. 1: Time-evolving anomaly detection performance of different approaches

Table 2: Information of datasets

	nodes	edges	attributes
Wiki	2,405	10,976	4,973
Blogcata	4,654	148,372	8,189
Flickr	7,000	203,834	12,047

Table 3: Average running time

	ANOMALOUS	AMAD
Wiki	579.20(s)	**147.30(s)**
Blogcata	885.31(s)	**686.92(s)**
Flickr	5136.81(s)	**2426.04(s)**

4 Conclusions

In this paper, we propose a novel dynamic anomaly detection framework AMAD and experiments corroborate the effectiveness of AMAD. Additionally, future work can be focused on detecting group anomaly in a dynamic setting.

Acknowledgements. This work is supported by National Key Research and Development Program of China (2016YFB1000903), National Nature Science Foundation of China (61872287, 61532015 and 61672418), Innovative Research Group of the National Natural Science Foundation of China (61721002), Innovation Research Team of Ministry of Education (IRT_17R86), Project of China Knowledge Center for Engineering Science and Technology.

References

1. Aggarwal, C., Subbian, K.: Evolutionary network analysis: A survey. ACM Computing Surveys (CSUR) **47**(1), 10 (2014)
2. Akoglu, L., Tong, H., Vreeken, J., Faloutsos, C.: Fast and reliable anomaly detection in categorical data. In: CIKM (2012)
3. Breunig, M.M., Kriegel, H.P., Ng, R.T., Sander, J.: Lof: identifying density-based local outliers. In: ACM sigmod record. vol. 29, pp. 93–104. ACM (2000)
4. Li, J., Dani, H., Hu, X., Liu, H.: Radar: Residual analysis for anomaly detection in attributed networks. In: IJCAI (2017)
5. Peng, Z., Luo, M., Li, J., Liu, H., Zheng, Q.: Anomalous: A joint modeling approach for anomaly detection on attributed networks. In: IJCAI (2018)
6. Teng, X., Lin, Y.R., Wen, X.: Anomaly detection in dynamic networks using multi-view time-series hypersphere learning. In: CIKM (2017)