Source Locating of Spreading Dynamics in Temporal Networks*

Qiangjuan Huang
School of Science
National University of Defense Technology
Changsha, Hunan, China
qiangjuanhuang@foxmail.com
«Supervised by Professor Dongyun Yi, Chengli Zhao and Xue Zhang»

ABSTRACT

The topological structure of many real networks changes with time. Thus, locating the sources of a temporal network is a creative and challenging problem, as the enormous size of many real networks makes it unfeasible to observe the state of all nodes. In this paper, we propose an algorithm to solve this problem, named the backward temporal diffusion process. The proposed algorithm calculates the shortest temporal distance to locate the transmission source. We assume that the spreading process can be modeled as a simple diffusion process and by consensus dynamics. To improve the location accuracy, we also adopt four strategies to select which nodes should be observed by ranking their importance in the temporal network. Our paper proposes a highly accurate method for locating the source in temporal networks and is, to the best of our knowledge, a frontier work in this field. Moreover, our framework has important significance for controlling the transmission of diseases or rumors and formulating immediate immunization strategies.

Keywords

Source Locating; Temporal Network; Shortest Paths; Spreading Dynamics; Centrality

1. PROBLEM

Epidemic dynamical behavior can be observed in many real networks. Prototypical examples include epidemics transmission through social networks, virus propagation in communication networks, rumor transmission on the Internet, the cascading failure of power networks, and crises contagion through financial networks. Throughout history, epidemic outbreaks have brought about enormous losses for economies and society, from SARS in 2003 to the recent H7N9 flu outbreak. Therefore, many significant studies have examined the dynamics of epidemic outbreaks on networks [17, 16]. However, these studies have focused on the forward problem of the diffusion process and its dependence on the rate of infection. Another challenging question regarding the diffusion process is the inverse problem of inferring the original source in a huge network relying on relatively limited observed nodal states. This problem has received considerable attention in recent years [10, 3].

*(Produces the permission block, and copyright information). For use with SIG-ALTERNATE.CLS. Supported by ACM.

WWW'17 Companion, April 3-7, 2017, Perth, Australia.
ACM 978-1-4503-4913-0/17/04.
http://dx.doi.org/10.1145/3038912.3038914

2. STATE OF THE ART

The traditional analysis of source location is conducted on static networks, where it is assumed that the speed of a network's evolution is slower than the information transmission rate. In 2012, Pinto et al. established a likelihood function connecting the real infected time lags with theoretical time lags, enabling the source to be located from relatively few observations [18]. To locate the source, Brockmann and Helbing identified an effective transmission distance to calculate the spreading time between nodes [4], whereas Fabrizio et al. constructed a Bayesian conditional probability model of all nodal states, and found the source using marginal probabilities via belief propagation [1]. Actually, most real networks are dynamic and temporal. There are many methods for modeling such temporal networks. In 2012, Holme and Saramaki proposed the idea of using line graphs to describe temporal networks, before Nakamura and Tanizawa introduced a linear model of time-varying properties that could be transformed into a network structure [6, 13]. Recently, the special characteristics of temporal networks have attracted considerable attention. Some scholars have studied the influence of "burstiness" on the spreading process [11]. In 2011, Karsai et al. derived a Poisson distribution to describe a network, and proposed a transmission ratio between this Poisson distribution and a power-law distribution based on the SI (susceptible-infected) model [8]. Similarly, a number of researchers have investigated other characteristics of temporal networks that have an impact on information propagation [12]. However, research into locating the source of the spreading process in temporal networks is inadequate. To this end, Liu et al. recommended an algorithm for optimizing particle swarms with scale-free interactions to find the optimum or hub nodes in the evolution process [9]. And Antulov-Fantulin et al. proposed a statistic method for finding the spreading source based on the time information of all nodes is known [2].

3. PROPOSED APPROACH

In this paper, we focus on depicting the spreading process and locating the transmission source in a temporal network. To the best of our knowledge, this is the first work in this field. We propose a diffusion process and consensus dynamics to model the spreading process based on the definition of a temporal network. We then propose a backward temporal diffusion process (BTDP) to effectively locate the spreading source in the temporal network. We also improve the strategy of selecting the observed nodes by expanding the indexes that measure the nodes' importance to the temporal network, such as the degree, closeness centrality, and PageRank. Our framework for locating the source in a temporal network has showed a high accuracy and wide applications. We believe this work has considerable significance to the problem of source location, and has the potential to drive the study of temporal networks. However, there are also some problem needed to be resolved, like effectively calculating the shortest temporal paths in temporal network and applying our method of source locating into real networks.

4. METHODOLOGY

Generally, a network can be expressed by a binary group $G = (V, E)$, where V is the set of all nodes and E is the set of all links. Thus, we can also use $G = (V, E)$ to represent temporal networks, where V is still the set of nodes, but E is the set of events $e_{uv} = (u, v, t_{uv}, \Delta t_{uv})$, which means that nodes u and v interact at time t_{uv} for a period of Δt_{uv}.

4.1 Depicting the Spreading Process in Temporal Networks

Any spreading process, like epidemic spreading in a population, virus propagation on the Internet, rumor propagation in social networks and risk contagion in financial networks, can be regarded as a diffusion process. To be as general as possible, we consider a simple diffusion model associated with diffusion delay. In temporal networks, we assume that the diffusion process begins from a source node, and the information takes some time to reach its neighbor nodes because of the propagation delay along the links. As the network structure is not constant, each node that receives the information forwards it to its neighbors at that moment. The information transmits through effective temporal paths composed of several events $e_{v_0 v_1}, e_{v_1 v_2}, \cdots, e_{v_{n-2} v_{n-1}}, e_{v_{n-1} v_n}$, satisfying the condition $t_{e_{v_{i-1} v_i}} \geq t_{e_{v_{i-2} v_{i-1}}} + \Delta t_{e_{v_{i-2} v_{i-1}}}$, $\forall 1 \leq i \leq n$. The transmission continues until all nodes in the network have received the information.

Another method of modeling the dynamical spreading process is consensus dynamics, in which the aim is to reach an agreement regarding a certain quantity of interest that depends on the state of all agents. And consensus dynamics on complex networks have been investigated since the development of complex network science a decade ago. In most real systems, agreement and synchronization phenomena are similar to the consensus of linear systems to some extent, and we know the nodes state in spreading processes will reach to a steady state, so we can use the consensus dynamics to model the spreading process [14]. In a static network, we can use the consensus algorithm under communication time delays to model the spreading process, $\dot{x_i} = \sum_{j=1}^{N} a_{ij}[x_j(t - \tau_{ij}) - x_i(t)]$. But in temporal network the network's structure changes with time, the adjacency matrix can be defined as $A(t) = (a_{ij}(t))_{N \times N}$, and the consensus dynamics are given by:

$$\dot{x_i} = \sum_{j=1}^{N} a_{ij}(t)[x_j(t - \tau_{ij}) - x_i(t)] \tag{1}$$

4.2 The BTDP Model for Source Locating

In this paper, we put forward a backward diffusion method by calculating the shortest temporal distance between nodes to locate the propagation source in temporal networks, which is inspired from the work of Shen, who proposed an efficient method called time reversal virtual diffusion to locate the source with high accuracy in static network [19]. Because, in temporal networks, the structure is not constant and the shortest paths between nodes are limited with respect to time, so the source can be located by analyzing the structural characteristics of the temporal network. We assume that the temporal network structure and the time information of some observed nodes are known in advance. In a real network, the propagation delay along a link between nodes is ambiguous because of the perturbations caused by various random factors. However, we can assume that the time delay follows a certain distribution, namely a Gaussian or uniform distribution. The mean value and variance of these two distributions are finite, and are easily observed. Thus, we propose the backward temporal diffusion process model to locate the source in a temporal network. The BTDP algorithm contains two steps:

1. Backward the diffusion process from observed nodes $\{o_1, o_2, \cdots, o_m\}$. Because the propagation times $\{t_{o_1}, t_{o_2}, \cdots, t_{o_m}\}$ of the observed nodes are known, we can calculate the shortest time from any node i to the observed nodes o_k, denoted as $t(i, o_k)$. Thus, we obtain a vector $\mathbf{T}_i = [t_{o_1} - t(i, o_1), t_{o_2} - t(i, o_2), \cdots, t_{o_m} - t(i, o_m)]^{\mathbf{T}}$ for every node. The problem is to determine the shortest path between two nodes in the temporal network.

2. Calculating the variance of the vectors $\{\mathbf{T}_1, \mathbf{T}_2, \cdots, \mathbf{T}_N\}$. The node with the minimum variance is the source.

The key and difficult part of this method lies in calculating the shortest temporal distance between every pair of nodes.We now put forward a sample method based on defining the temporal path to calculate the shortest temporal distance. As the temporal path can be expressed as a series of events $e_{v_0 v_1}, e_{v_1 v_2}, \cdots, e_{v_{n-2} v_{n-1}}, e_{v_{n-1} v_n}$. If there is a temporal path from v_0 to v_n, marked as P, consisted with $e_{v_0 v_1} = (v_0, v_1, t_1, \Delta t_1), e_{v_1 v_2} = (v_1, v_2, t_2, \Delta t_2), \cdots, e_{v_{n-1} v_n} = (v_{n-1}, v_n, t_n, \Delta t_n)$, we can define the temporal distance as $dist(v_0, v_n) = t_n + \Delta t_n - t_0$, and the shortest temporal distance is the shortest among all the temporal paths from v_0 to v_n we can fine, $mindist(v_0, v_n)$. Here we proposed a sample algorithm to find the shortest temporal distance between a pair of nodes with describing the temporal network as edge sequence, sorting as the time order.

Algorithm 1: Computing the shortest temporal distance

Input: A temporal graph $G = (V, E)$ in its edges sequence presentation, source vertex s

Step1: Initialize $t_s = t_0$ and $t_v = \infty$ for all $v \in V \setminus n$

Step2: for each incoming edge $e = (u, v, t_{uv}, \Delta t_{uv})$ form the edge sequence

    if $t_{uv} \geq t_u$
        if $t_{uv} + \Delta t_{uv} - t_0 < t_v$ then
            $t_v \leftarrow t_{uv} + \Delta t_{uv} - t_0$
    else
        break the for-loop.

Output: The shortest temporal distance from s to every vertex $v \in V \setminus n$.

4.3 Observer Selection Strategies

In the above source location method, we have the hypothesis that the nodes with known propagation times are observed at random. As the network structure is known, we can select the observers by measuring the importance of nodes. The analysis of a node's importance is also a hot issue in the field of complex networks, and can be used to find the key nodes in the spreading process [7]. In a static network, common indexes of nodes' importance include the degree, closeness centrality [20], k-shell [5], and PageRank [15]. In a temporal network, we can redefine these indexes to measure the nodes' importance. And we replace the random selection strategy of obtaining the observed nodes in the first step of backward temporal diffusion process method with the different kinds of selection strategies based on the nodes' importance measurements.

Degree: As the structure of a temporal network changes with time, the node degrees also change. Thus, we obtain a series of degrees for every node, that is, $\{De_{i1}, De_{i2}, \cdots, De_{iT}\}$. To determine a node's degree, we can sum the series of degrees, $De_i = \sum_{t=1}^{T} De_{it}$ or choose the maximum, $De_i = max|De_{it}|$. In the proposed BTDP model, we use the latter method.

Closeness centrality: This index measures the distance of a node with respect to the other nodes. In a static network, it is defined as: $C_i = \frac{N-1}{\sum_j d_{ij}}$ where d_{ij} is the shortest distance between the node i and the node j. In the previous section, we defined the shortest temporal distance in a temporal network. Similarly, the closeness centrality can be defined as: $C_i = \frac{N-1}{\sum_j \tau_{ij}}$

PageRank number:The idea is that the importance of one node depends on the neighbor nodes directed towards it and the value of these nodes. First, we set an initial PR_i for every node, and then the PageRank number of every node at step $k-1$ is evenly distributed to all neighbor nodes. The iteration stops when the PageRank number of every node is stable. Thus, the iteration equation is: $PR_i(k) = \sum_{j=1}^{N} a_{ji} \frac{PR_j(k-1)}{d_j^{out}}$. In a temporal network, we obtain PageRank numbers for every node at different times and we also choose the maximum PageRank number among the time series, $PR_i = max|PR_{it}|$.

5. RESULTS

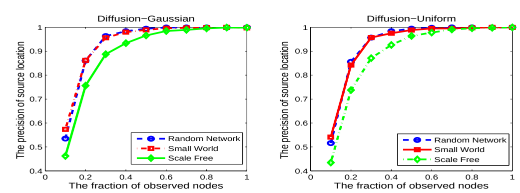
In order to analyze the precision of the model for locating the source, we simulate three classic types of networks, including random network(ER), small-world network(WS) and scale-free network(BA), $N = 100$ and $\langle k \rangle = 8$, $T = 10$. The time delays of the links were assumed to follow a Gaussian distribution with mean 1.0 and standard variance 0.25 or a uniform distribution in the range (0.5, 1.5).

5.1 The Precision of Location

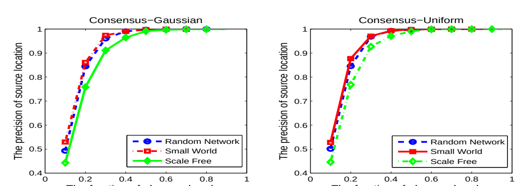
From Fig. 1, we can see that the method of backward diffusion with temporal shortest path performs best in random networks and worst in scale-free networks, regardless of spreading model and time delay distribution. Under the simple diffusion model, if we require 90% precision for locating the source (i.e., in 100 independent experiments, we can find the source 90 times), we need to observe 23%, 23%, and 34% of nodes when the time delays obey a Gaussian distribution and 24%, 24%, and 35% of nodes when the time delays obey a uniform distribution in the ER, WS, and BA networks. Thus, the distribution type has little effect on the precision of source location. For the consensus dynamics spreading process, the proportion of nodes that must be observed for 90% precision is 24%/23%, 23%/21%, 29%/30% for the Gaussian and uniform distributions, show in Table. 1. Clearly, the performance of source locating method shows similar to each other in Random networks and in small-world networks, and more observed nodes needed in scale-free networks if we want to get the same locating precision.

Table 1: The percentage of observed nodes needed when the locating precision reach to 90%

Network	ER		WS		BA	
Distribution	Gau	Uni	Gau	Uni	Gau	Uni
Spreading	23%	24%	23%	24%	34%	35%
Consensus	24%	23%	23%	21%	29%	30%



(a) Diffusion Process



(b) Consensus Dynamics

Figure 1: Comparison of location precision in random, small-world, and scale-free networks when the time delay obeys a Gaussian or uniform distribution

5.2 Influence of Selection Strategies

If we selected the observed nodes according to the rank of the nodes' importance, rather than randomly. The influence of these three strategies on the location precision can be seen in Fig. 2 and Fig. 3.We can see that the strategies of selecting which nodes to observe can affect the locating precision in some extent, in the scale-free networks. It is known that the majority of real-world networks, such as social and biological networks, are scale-free. Tab. 2shows the percentage of observed nodes needed when the locating precision reach to 90% in the case that the time delay obeys a Gaussian distribution and the spreading is modeled as a diffusion process. The results show that source location is most effective if we choose the observed nodes based on their closeness centrality in any types of networks. And we can see that in scale-free network, if we want to obtain 90% precision of source locating, the observed nodes needed decreased from 34% to 24% for the diffusion process and form 29% to 18% for consensus dynamics respectively with applying the closeness centrality selection strategy when the time delay obey the Gaussian distribution, which is a great obvious improvement in the large scale networks.

Table 2: The percentage of observed nodes needed when the locating precision reach to 90% in the case that the time delay obeys a Gaussian distribution and the spreading is modeled as a diffusion process.

Strategies	Random	Degree	Closeness	PageRank
ER	23%	21%	18%	20%
WS	23%	26%	21%	27%
ER	34%	24%	24%	26%

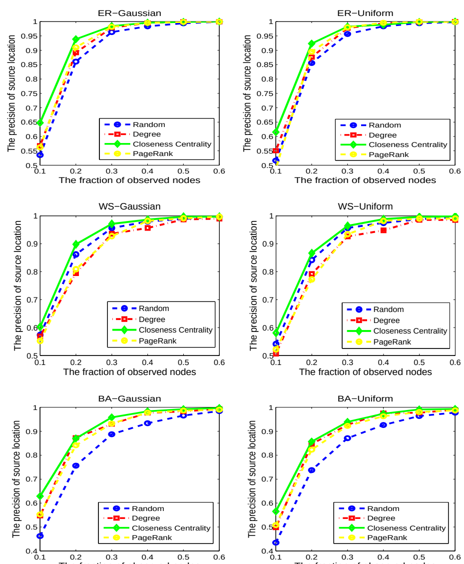


Figure 2: Location precision with different selection strategies when the time delay obeys a Gaussian or uniform distribution and the spreading is modeled as a diffusion process.

5.3 The Influence of Parameters

In the BTDP source location method, two parameters should be considered: the average degree, which measures the sparseness of the network, and the length of the time series describing the network's structural evolution. Thus, we simulated three network types with average degrees of 6, 8, and 10. From Fig. 4, we can see that a higher average degree increases the precision of source location for the same proportion of observed nodes. A higher degree implies that

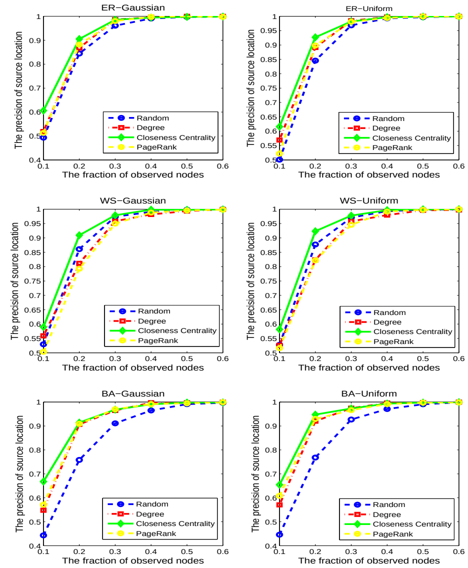


Figure 3: Location precision with different selection strategies when the time delay obeys a Gaussian or uniform distribution and the spreading is modeled by consensus dynamics.

the network is denser, meaning that the temporal distance between nodes is shorter. This reduces the estimated deviation in time delay in the location process. To analyze the impact of the time series length, we simulated temporal networks with lengths of 10, 15, and 20. From Fig. 5, we can see that the precision changes very little for different lengths of temporal network. We believe this is because the temporal shortest distance is always less than the time series length, so the length has little influence on our source location model. Through the above analysis, we can conclude that our location model is stable and practical in real networks.

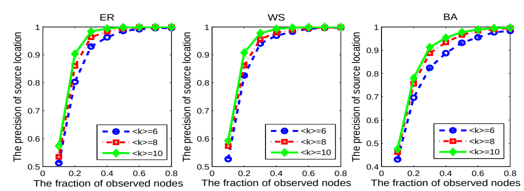


Figure 4: Influence of the network's average degree on the location precision of three types networks

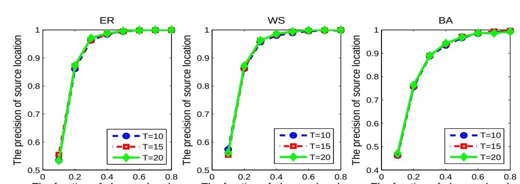


Figure 5: Influence of the time series length on the location precision of three types networks

6. CONCLUSIONS AND FUTURE WORK

As the concept of temporal networks has become more widespread, researchers have begun to study their topological characteristics and complex dynamical processes, such as link prediction, community discovery, and transmission dynamics. In this paper, we have focused on the inverse problem of source location in a temporal network. We first introduced the concept of temporal networks and defined some of their structural properties concerning events and effective paths. We then proposed a simple diffusion model and consensus dynamics to depict the spreading process within a temporal network. For source location, we proposed the BTDP method based on the calculation of temporal shortest paths. The strategy of observing a random set of nodes was found to be an inefficient means of locating the source. Thus, we extended several indexes of node importance in static networks to temporal networks, and chose which nodes to observe according to the rank of these indexes. To some extent, these strategies improved the location precision. Finally, we examined the average degree and time series length, which verified the stability and practicality of the proposed location model.

Some aspects of source location modeling require further discussion. First, in our source location method, the computation of shortest temporal distance need to traverse all the nodes in different time period which is much complex to the large scale networks. Thus, it is important to develop new methods to calculate the temporal shortest paths or use other measurements of temporal distance. Second, in this paper, we used both a simple diffusion process and consensus dynamics to simulate the information spreading process. However, classical methods such as SI, SIS, and SIR use ordinary differential equations. Combining temporal information with these models to describe the spreading process and locate the source is an important area of future study. Finally, real-world temporal networks often have multiple sources. In practice, disease diffusion and rumor transmission does not always start from a single source. Whether the proposed location method remains useful in this multi-source scenario requires further investigation.

7. REFERENCES

[1] F. Altarelli, A. Braunstein, L. Dall'Asta, A. Lage-Castellanos, and R. Zecchina. Bayesian inference of epidemics on networks via belief propagation. *Physical Review Letters*, 112(11):118701–118701, 2014.

[2] N. Antulovfantulin, A. Lancic, S. T, S. H, and S. M. Identification of patient zero in static and temporal networks: Robustness and limitations. *Physical Review Letters*, 114(24), 2015.

[3] N. Antulovfantulin, A. Lancic, H. Tefancic, M. Ikic, and T. Muc. Statistical inference framework for source detection of contagion processes on arbitrary network structures. In *2014 IEEE Conference on Self-Adaptive and Self-Organizing Systems Workshops (SASOW)*, pages 78–83, 2013.

[4] D. Brockmann and D. Helbing. The hidden geometry of complex, network-driven contagion phenomena. *Science*, 342(6164):1337–42, 2013.

[5] S. Carmi, S. Havlin, S. Kirkpatrick, Y. Shavitt, and E. Shir. From the cover: A model of internet topology using k-shell decomposition. *Proceedings of the National Academy of Science*, 104(27):11150–11154, 2007.

[6] P. Holme and J. Saramaki. Temporal networks as a modeling framework. *Understanding Complex Systems*, pages 1–14, 2013.

[7] M. Kitsak, L. K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H. E. Stanley, and H. A. Makse. Identifying influential spreaders in complex networks. *Nature Physics*, 6(11):888–893, 2010.

[8] M. Kjaergaard, S. Brander, and F. M. Poulsen. Small but slow world: how network topology and burstiness slow down spreading. *Physical Review E*, 83(83):602–608, 2010.

[9] C. Liu, W. B. Du, and W. X. Wang. Particle swarm optimization with scale-free interactions. *Plos One*, 9(5):399–402, 2014.

[10] A. Y. Lokhov, M. MÍęzard, H. Ohta, and L. ZdeborovÍć. Inferring the origin of an epidemic with a dynamic message-passing algorithm. *Physical Review E Statistical Nonlinear and Soft Matter Physics*, 90(90):012801–012801, 2014.

[11] B. Min and K. Goh. *Burstiness: Measures, Models, and Dynamic Consequences*. Springer Berlin Heidelberg, 2013.

[12] B. Min, K. I. Goh, and A. Vazquez. Spreading dynamics following bursty human activity patterns. *Physical Review E Statistical Nonlinear and Soft Matter Physics*, 83(3 Pt 2):515–529, 2011.

[13] T. Nakamura and T. Tanizawa. Networks with time structure from time series. *Physica A Statistical Mechanics and Its Applications*, 391(20):4704–4710, 2012.

[14] R. Olfati-Saber, A. Fax, and R. M. Murray. Consensus and cooperation in networked multi-agent systems. *Proceedings of the IEEE*, 95(1):215–233, 2007.

[15] L. Page. The pagerank citation ranking : Bringing order to the web. 9(1):1–14, 1998.

[16] R. Pastor Satorras and V. Alessandro. Epidemics and immunization in scale-free networks. *Bornholdt S and Schuster H G Handbook of Graph and Networks*, pages 111–130, 2002.

[17] R. Pastor Satorras and A. Vespignani. Epidemic dynamics and endemic states in complex networks. *Physical Review E Statistical Nonlinear and Soft Matter Physics*, 63(6 Pt 2):138–158, 2001.

[18] P. C. Pinto, P. Thiran, and M. Vetterli. Locating the source of diffusion in large-scale networks. *Phys.rev.lett*, 109(6):1–5, 2012.

[19] Z. Shen, S. Cao, Y. Fan, Z. Di, W. X. Wang, and H. E. Stanley. Locating the source of spreading in complex networks. *Computer Science*, 2015.

[20] J. Tang, S. Scellato, M. Musolesi, C. Mascolo, and V. Latora. Small-world behavior in time-varying graphs. *Physical Review E Statistical Nonlinear and Soft Matter Physics*, 81(2):055101–055101, 2010.