Physica A 451 (2016) 1–9

Contents lists available at ScienceDirect

Physica A

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

# Explosive percolation in thresholded networks

Satoru Hayasaka

*Department of Psychology, The University of Texas at Austin, Austin, TX, 78712, USA*
*Department of Biostatistical Sciences, Wake Forest University Health Sciences, Winston–Salem, NC, 27157, USA*
*Department of Radiology, Wake Forest University Health Sciences, Winston–Salem, NC, 27157, USA*

## HIGHLIGHTS

- Explosive percolation can be observed in real networks formed by thresholding.
- Explosive percolation is possible when edges are added in a particular order.
- The proposed method does not involve a random process.
- Examples of explosive percolation with real network data are shown.

## ARTICLE INFO

*Article history:*
Received 7 July 2015
Received in revised form 12 October 2015
Available online 28 January 2016

*Keywords:*
Explosive percolation
Networks
Correlation matrix
Phase transition

## ABSTRACT

Explosive percolation in a network is a phase transition where a large portion of nodes becomes connected with an addition of a small number of edges. Although extensively studied in random network models and reconstructed real networks, explosive percolation has not been observed in a more realistic scenario where a network is generated by thresholding a similarity matrix describing between-node associations. In this report, I examine construction schemes of such thresholded networks, and demonstrate that explosive percolation can be observed by introducing edges in a particular order.


## 1. Introduction

Percolation is a phase transition phenomenon where a unique large connected cluster emerges in a lattice or a network, as connections are gradually introduced. A well-known example of percolation in networks is the Erdős–Rényi (ER) model [1], in which $n$ isolated nodes are randomly connected by $m$ edges. Here, the relationship between the number of edges and nodes can be described by the fraction $t$ defined as $t = m/n$, or $m = tn$. When $t < 1/2$ (or $m < n/2$), the size of the largest connected component (known as the giant component) $S_{\max}$ is small. However, at $t = 1/2$, a unique giant component emerges covering a large portion of the network. As $t$ increases further ($t > 1/2$), the giant component grows until it encompasses all the available nodes. A number of recent papers examine how such percolation can be predicted in various types of networks [2–5]. The number of steps required for such percolation can be shortened by simple schemes, giving an appearance of an abrupt phase transition known as explosive percolation [6]. A number of methods, modified versions of the ER model, have been reported to produce explosive percolation [7,8]. Explosive percolation can be observed not only on ER networks, but also in other types of random network models [9,10]. A recent review on explosive percolation can be found in Ref. [11]. Explosive percolation schemes involve adding edges randomly in a manner that prevents formation of large clusters. Such process can delay percolation [6] while setting up a collection of connected components, known as the *powder keg* [7,8], capable of producing explosive percolation. However, such schemes also introduce randomness since edges are

*E-mail address:* hayasaka@utexas.edu.

http://dx.doi.org/10.1016/j.physa.2016.01.001
0378-4371/

added in a random fashion. Consequently such schemes are only relevant to random network models, as the same network cannot be reproduced again. Existing real networks can be *reconstructed* to exhibit explosive percolation by applying such a scheme [12,13]. However, it is not clear if explosive percolation can be observed during the construction of a real network without introducing a stochastic process commonly seen in the existing explosive percolation schemes. In this report, I present a simple algorithm to construct a network during which explosive percolation is observed. In particular, I focus on a class of networks that can be constructed by thresholding a similarity matrix describing the strength of associations between nodes (e.g., a correlation matrix).

## 2. Methods and materials

### 2.1. Thresholded networks

In a similarity matrix, each row or column represents a node in the network, and the *ij*-th element quantifies the association between nodes $i$ and $j$ (see Fig. 1(a)). If the *ij*-th element exceeds a certain threshold, then nodes $i$ and $j$ are considered connected by an edge. The resulting network is often an undirected network. Networks constructed by thresholding a similarity matrix, or thresholded networks, are often products of hard thresholding, in which the same threshold value is applied for the entire matrix. However, hard thresholding often leads to concentration of edges in some parts of networks while a large portion of nodes may be disconnected from the rest of the network [14,15]. One way to overcome this problem is to threshold each row of the similarity matrix separately, controlling the number of edges originating from the corresponding node [14,15]. Another way to overcome the problem is to only retain edges with statistically significant weights at each node [16,17]. These methods are known to preserve the backbone of the underlying complex network [14,16,17]. In this report, I adopt the thresholding method proposed by Ruan et al. [15], referred as rank-based thresholding, in which the top $d$ highest values are identified in each row of a similarity matrix and the corresponding edges are added to the network. Even with a small value of $d$ ($\simeq 3$), the resulting network is likely to be connected [15].

Although Ruan et al.'s method can produce a connected network with a relatively small number of edges [15], they did not examine how a network evolves as edges are added to isolated nodes one at a time, in a similar manner as the construction of a random network model. Since rank-based thresholding can produce a connected graph, percolation may be observed as edges are added one-by-one, depending on the order edges are added. Moreover, such percolation may be explosive if an appropriate scheme is chosen to add edges. To this end, I examined two approaches of constructing a thresholded network. In both approaches, the largest elements (largest, 2nd largest, 3rd largest, …, up to $d$th largest) were identified in each row of the similarity matrix (Fig. 1(b)). Each of these elements represented an edge $(i, j)$, where $i$ and $j$ were row and column indices, respectively (Fig. 1(b)). Then these largest elements were sorted within each rank, then concatenated as shown in Fig. 1(c). The elements can be sorted in ascending order within each rank; I shall refer this as the *ascending approach*. Or, the elements can be sorted in descending order within each rank, referred as the *descending approach*. The sorted edges were added to the network, one-by-one, in the order in the concatenated vector (see Fig. 1(c)). It should be noted that, in both approaches, the same edge may be selected twice (e.g., edges (1, 10) and (10, 1) in Fig. 1(c)). If that occurred, then only the first edge was added to the network while the second edge was discarded. In a network generated by rank-based thresholding, node degrees were not $d$ for all the nodes. Even if an edge $(i, j)$ is attributed to one of the top $d$ values for node $i$, it may not be part of the top $d$ values for node $j$. This results in node degree of $j$ greater than $d$.

### 2.2. Network data

Four examples of thresholded networks were examined, namely, a stock market network, an airline passenger traffic network, a gene co-expression network, and brain functional connectivity networks.

#### 2.2.1. Stock market network

The stock price data of 491 companies listed in Standard & Poor's 500 index (S&P500) were downloaded using the get.hist.quote function in the tseries package of R. In particular, the adjusted closing prices between January 1, 2000 and December 31, 2013 were downloaded for these companies. The downloaded time series data were then converted to the one-period fractional return $r(t) = (p(t) - p(t-1))/p(t-1)$ where $p(t)$ is the stock price at the time point $t$. Then the correlation coefficients between $r(t)$'s from different companies were calculated to generate a correlation matrix. Only the time points where both companies had the price data were used in the calculation in the correlation coefficients; the number of time points varied between 47 and 3467. The calculation resulted in a $491 \times 491$ correlation matrix.

#### 2.2.2. Airline passenger traffic network

The US domestic airline passenger traffic data for year 2013 were downloaded from the Bureau of Transportation Statistics from the United States Department of Transportation.[1] The data listed the number of passengers from one airport

[1] www.rita.dot.gov/bts/.