# A Two-Parameter Method to Characterize the Network Reliability for Diffusive Processes

Madhurima Nath[1,2], Stephen Eubank[1,3], Mina Youssef[1,⋆], Yasamin Khorramzadeh[1,2], and Shahir Mowlaei[1,2]

[1] Network Dynamics and Simulation Science Laboratory, Virginia Bioinformatics Institute, Virginia Tech, Blacksburg, Virginia 24061, USA
[2] Department of Physics, Virginia Tech, Blacksburg, Virginia 24061, USA
[3] Department of Population Health Sciences, Virginia Tech, Blacksburg, Virginia 24061, USA
{nmaddy,seubank,myoussef,yasi,shahir}@vbi.vt.edu

**Abstract.** We introduce a new method to characterize the network reliability polynomial of graphs – and hence the graph itself – using only a few parameters. Exact evaluation of the reliability polynomial is almost impossible for large graphs; estimating the polynomial's coefficients is feasible but requires significant computation. Furthermore, the information required to specify the polynomial scales with the size of the graph. Thus, we aim to develop a way to characterize the polynomial well with as few parameters as possible. We show that the error function provides a two-parameter family of functions that can closely reproduce reliability polynomials of both random graphs and synthetic social networks. These parameter values can be used as statistics for characterizing the structure of entire networks in ways that are sensitive to dynamical properties of interest.

**Keywords:** Network reliability, Error function, synthetic social networks.

## 1 Introduction

### 1.1 Motivation

It has been more than 50 years since Moore and Shannon introduced the network reliability polynomial to study the performance of electronic circuits with "crummy" relays [1]. Since then, the concept has been widely applied in designing reliable circuits and other networks delivering commodities between source and destination locations. Early studies showed the effect of network topology on the overall performance of the network for simple commodity flow between a source vertex and a sink vertex [2,3]. This problem is well-known as the *Two-Terminal reliability rule.* Another common performance measure is the probability that a

⋆ Corresponding author.



G. Mangioni et al. (eds.), *Complex Networks VI,*
Studies in Computational Intelligence 597, DOI: 10.1007/978-3-319-16112-9_14

randomly selected set of edges connects all the vertices of the original graph, often referred to as the *All-Terminal reliability rule*. Furthermore, the *K-Terminal reliability rule* – the probability that a randomly selected subgraph contains a predefined set of vertices of size $K$ – has been also studied. For a comprehensive review of the reliability polynomial, we refer the reader to the book by C. J. Colbourn [4].

Network reliability polynomials are not limited to Two-Terminal reliability, $K$-Terminal reliability or All-Terminal reliability rules. There are many features of percolation processes that the network reliability polynomial can reflect. In previous studies, the concept of reliability polynomial was successfully applied to study the spread of the infectious diseases in social networks [5,6].

The classical concept of network reliability provides a rich theoretical basis, supported by computational estimation procedures, to study the effect of structural properties on the diffusion of dynamics. Although evaluation of the reliability polynomial coefficients is usually intractable and its complexity is #-P hard [4], estimating the coefficients to within a practically important confidence interval is feasible. This paper aims at shrinking the wide gap between theoretical analysis of reliability problems and our ability to apply the conceptual framework to practical problems for large and non-trivial graphs. The estimation procedure relies on the random selection of subgraphs from the main graph under study. The reliability rule, which is chosen based on the dynamical features of interest, is applied to every subgraph to determine whether it exhibits the desired feature. For example, in this paper, the feature we are interested in is the probability that a certain fraction $\alpha$ (the "attack rate")of the population will be infected during an outbreak of disease, as a function of its person-to-person transmissibility.

The reliability polynomial describes the system's behavior. We would like to use it to characterize the system itself, and to that end, we need a way to summarize the information it contains. We can take advantage of the fact that, for rules that satisfy a simple criterion, the reliability polynomials are monotonic increasing sigmoidal functions from the interval $[0, 1]$ to itself. This suggests representing the polynomial as the cumulative distribution function of a continuous probability density.

### 1.2 Contribution

In this paper, we represent the network reliability polynomial in terms of well-known two-parameter functions. Here, we test two functions, the error function and the binomial cumulative distribution function. We fit reliability polynomials for several random graphs and synthetic social networks to these functions, note the values of the best-fit parameters, and evaluate the goodness of fit using the statistical coefficient of determination. The error function provides better fits than the binomial CDF fits and provides a very close fit to all the examples. Thus the two parameters of the error function, the mean and deviation of the corresponding Gaussian, were sufficient to characterize the reliability polynomial.

We observe that, for random graphs, the values of these two parameters are weakly correlated with the size of the graph. The values also depend on the reliability rule. Lastly we exploit this method to characterize differences among synthetic social networks for the New River Valley in Virginia, Mexico City, Sierra Leone and Liberia [7]. We conclude that we can reconstruct the network's reliability using just a few parameters.

The paper is organized as follows: Section 2 introduces the definition of network reliability and reliability rules. Section 3 elaborates on fitting the reliability polynomial to two parameterized functions. The numerical evaluations are described in Section 4. Finally, the conclusions are discussed in Section 5.

## 2 Network Reliability Polynomial

Moore and Shannon introduced the concept of the network reliability polynomial in the 1950's to evaluate the performance of electrical circuits composed of crummy relays. Given that every relay has a probability of failure, Moore and Shannon showed that the probability the circuit functioned as desired could be expressed as a polynomial. In addition, they evaluated the circuit reliability given that the relays are connected in series, in parallel, and in certain combinations of series and parallel. In this paper, we use the reliability concept to analyze social networks. In particular, we tie the concept of network reliability to network epidemiology by evaluating the probability of obtaining a given attack rate as a function of transmissibility.

### 2.1 Mathematical Definition

Given a graph composed of $N$ vertices and $E$ edges and a criterion that clearly defines the acceptance or the rejection of a subgraph–represented as the *reliability rule r:* $r(g) \in \{0, 1\}$–in a binary form, we introduce a *damage model* that assigns a probability to each subgraph. The network reliability is then:

$$R_G(\boldsymbol{x}) \equiv \sum_{g \subset G} r(g) p_{\boldsymbol{x}}(g) \tag{1}$$

where $g$ is a subgraph, $r(g)$ is 1 if the subgraph $g$ is accepted by the rule $r$, and $p_{\boldsymbol{x}}(g)$ is the probability to obtain the subgraph $g$ under the damage model. In this paper, the damage model includes each edge with probability $x$, corresponding to bond percolation. The probability of obtaining a subgraph with $k$ edges is $x^k(1-x)^{E-k}$. We denote the number of subgraphs with $k$ edges that are accepted by the reliability rule $r(g)$ as $R_k$. The network reliability can be written as a polynomial:

$$R_G(x) = \sum_{k=0}^{k=E} R_k x^k (1 - x)^{E-k}. \tag{2}$$

The term $R_k$ is called the reliability coefficient, and its computation is, in general, intractable. However, we know that there are $\binom{E}{k}$ subgraphs with $k$

edges in the graph, and that some fraction between 0 and 1 of them is accepted by the reliability rule. Therefore, the reliability coefficient can be written as follows:

$$R_k = P_k \binom{E}{k} \tag{3}$$

where $P_k$ is the fraction of with $k$ edges that is accepted. Hence:

$$R_G(x) = \sum_{k=0}^{k=E} P_k \binom{E}{k} x^k (1-x)^{E-k}. \tag{4}$$

Computing the coefficients $P_k$ is straightforward. Simply select a subgraph with $k$ edges randomly and evaluate the reliability rule. The estimate for $P_k$ is the number of accepted subgraphs divided by the total number of subgraphs sampled. The random selection of subgraphs is repeated until the sampling error for $P_k$ is within the desired confidence interval. The number of edges in the smallest accepted subgraph is called $k_{min}$, while the number of edges in the largest unaccepted subgraph is called $k_{max}$ [8]. Thus the $P_k$ curve has values between 0 and 1 between $k_{min}$ and $k_{max}$, respectively. The reliability $R(x)$ is a smoothed version of $P_k$, as can be seen from Equation 4, where $\binom{E}{k} x^k (1-x)^{E-k}$ plays the role of a sharply peaked smoothing kernel.

We use the following reliability rule: *a graph $g$ is accepted if and only if the mean square size of connected components in $g$ is greater than $\alpha N$*. This rule creates a mapping between an interesting epidemiological problem and the reliability polynomial as follows: $x$ represents the transmission probability, $\alpha$ represents the attack rate, and $R(x)$ represents the probability that the attack rate is at least $\alpha$. We denote this rule as $ExpX - \alpha$.

The motivation of this work was to characterize the reliability polynomial $R(x)$ using a small set of parameters. Since $R(x)$ has the properties of a cumulative distribution function (CDF), we propose to fit the $P_k$ values to the binomial CDF and the error function.

# 3 Two-Parameter Characterization of Network Reliability

## 3.1 Binomial CDF Method

The cumulative distribution function (CDF) of a binomial distribution is given as

$$CDF(k) = \sum_{m=0}^{k} \binom{N}{m} p^m (1-p)^{N-m}. \tag{5}$$

$R(x)$ has a similar form (Eqn. 4). Thus, we fit the right hand side of Eqn.(5) to the $P_k$ of Eqn.(4) and plot it with respect to the $k$ estimates $P_k$ by taking $p$ as a parameter. They are truncated between the $k_{min}$ and $k_{max}$. The $k$ data points are re-scaled to give the values of $m$ such that it runs from 0 to $N$ as

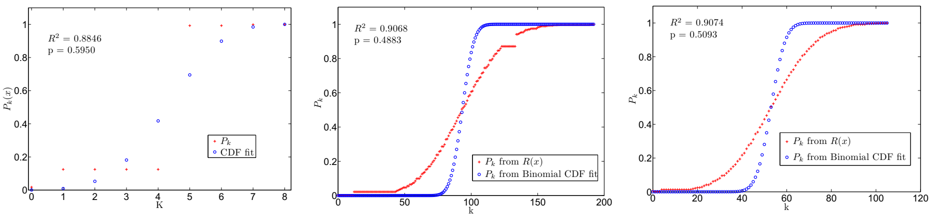


**Fig. 1.** Plot of $P_k$ and binomial CDF fit with $k^*$ for (a)$N = 20$, $M = 50$ (left panel), (b)$N = 2000$, $M = 50,000$ (middle panel) and (c)$N = 153,036$, $M = 4,152,739$ (right panel)

integers with an increment of 1. That is it runs from 0 to the total number of data points. Here, $N$ is chosen to be the maximum of $m$. We find the value of $p \in [0, 1]$ for which the binomial CDF best fits the $P_k$ values. We use the coefficient of determination $R^2$ as a measure of goodness-of-fit as follows:

$$R^2 = 1 - \sum_i \frac{(y_i - y_{fit})^2}{(y_i - y_{mean})^2} \tag{6}$$

where $y_i$ are the data points, $y_{mean}$ is the mean of the data points and $y_{fit}$ are the fitted data. The best fit is obtained when $R^2$ is closest to 1.

Figure 1 shows the binomial CDF fit to three sets of random networks generated by choosing a specific number of edges $M$ uniformly at random over a specific number of vertices $N$. These are (a)$N = 20$, $M = 50$, (b)$N = 2000$, $M = 50,000$ and (c)$N = 153036$, $M = 4,152,739$. As the size of the graph increases the fit becomes better as observed from the $R^2$ value, which increases from 0.8846 to 0.9074.

### 3.2 Error Function Method

We show the fit of $R(x)$ to the error function:

$$erf(X) = \frac{1}{a} \int_0^X e^{-(\frac{t+b}{a})^2} dt \tag{7}$$

Here, $a$ changes the width of the underlying Gaussian and is related to the variance whereas $b$ shifts the position of the mean of the Gaussian. The error function is defined between $-1$ and 1 for positive and negative values of $X$. The $R(x)$ and $P_k$ have values only between 0 and 1 and $x$ lies between 0 and 1. Thus, we rescale the error function such that both of them are in the same range. For this we fit $R(x)$ and $P_k$ to $\frac{1}{2}(erf(ax - b) + 1)$. Also, we normalize the $x$ values to $x^*$ given by

$$x^* = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{8}$$

**Table 1.** Comparison between fits to the error function method and the Binomial CDF

| Size of Graph | Two-parameter method | Binomial CDF method |
|---|---|---|
| $N = 20,\ M = 50$ | 0.9830 | 0.8846 |
| $N = 100,\ M = 450$ | 0.9508 | 0.8808 |
| $N = 500,\ M = 35750$ | 0.9976 | 0.8983 |
| $N = 2000,\ M = 50,000$ | 0.9993 | 0.9068 |
| $N = 10,000,\ M = 40,000$ | 0.9916 | 0.9627 |
| $N = 35,000,\ M = 750,000$ | 0.9975 | 0.9142 |
| $N = 125,000,\ M = 500,000$ | 0.9986 | 0.8843 |
| $N = 153,036,\ M = 4,152,739$ | 0.9999 | 0.9075 |

$R(x)$ shows a sharp transition when plotted against $x$ for large systems. This means that the values of $x_{min}$ and $x_{max}$ are closer for larger systems compared to a small size network. To look at the behaviour of $R(x)$ in the region of the sharp transition we re-scale the axis from $x$ to $x^*$.

## 4 Numerical Evaluation

We generate random Erdős-Rényi $G(N, M)$ graphs each having $N$ vertices and $M$ edges. We estimate the reliability polynomial for these graphs, and we use the coefficient of determination $R^2$ as a metric to determine the closeness of fitting the reliability curve using the error function and the Binomial CDF method. The graphs that have been used in this analysis are summarized in Table 1. The last GNM graph in the table is generated with the same number of vertices and edges as an estimated social contact for the New River Valley region near Blacksburg, Virginia. We use the reliability rule $ExpX = 0.2$. We use both $k$, the number of edges in the sub-graph and $x$, the ratio of $k$ to the total number of edges $E$ in our analysis. We also use the $k_{min}$ or alternatively $x_{min}$ and $k_{max}$ or $x_{max}$ values to normalize our data. Table 1 shows a comparison between the two-parameter method based on the error function and the Binomial CDF method for different graphs. Based on $R^2$ values, the error function is a better fit than the Binomial CDF. Therefore, in the rest of the numerical evaluation, we use the error function to represent the reliability polynomial.

An exhaustive search was done in the parameter space to find out the values of $a$ and $b$ with $R^2$ as a metric of goodness-of-fit for both $R(x)$ and $P_k$. Figures 2 and 3 show the fits to the error function for both $P_k$ and $R(x)$ for the GNM graphs.

We clearly observe that the parameters $a$ and $b$ decrease as the graph size increases. Meanwhile, the value of $R^2$ indicates that a better fit is obtained as the graph size increases for fitting $P_k$ and $R(x)$. Next, we change the reliability rule from $ExpX = 0.2$ to 0.4, 0.6, and 0.8 and repeat the analysis. The parameters $a$ and $b$ and $R^2$ are reported in Table 2.

We evaluate the reliability polynomials for the NRV, Mexico City, Liberia, and Sierra Leone synthetic social networks. The Sierra Leone and Liberia synthetic

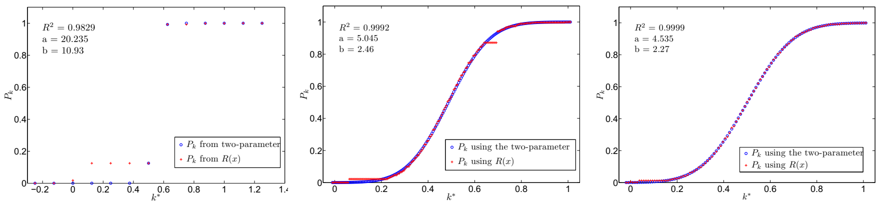


**Fig. 2.** Plot of $P_k$ and fit to the error function for GNM graphs with $k^*$ for (a)$N = 20$, $M = 50$ (left panel), (b)$N = 2000$, $M = 50,000$ (middle panel) and (c)$N = 153,036$, $M = 4,152,739$ (right panel)

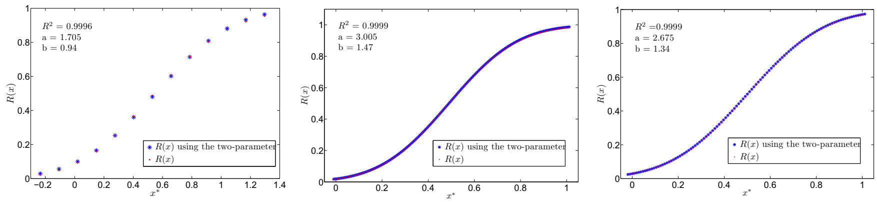


**Fig. 3.** Plot of $R(x)$ and the fit to the error function for GNM graphs with $x^*$ for (a)$N = 20$, $M = 50$ (left panel), (b)$N = 2000$, $M = 50,000$ (middle panel) and (c)$N = 153,036$, $M = 4,152,739$ (right panel)

**Table 2.** Fitting the reliability polynomial $R(x)$ of GNM graphs for different reliability rule $ExpX - \alpha$

| Graph | $\alpha = 0.4$: $R^2$,a,b | $\alpha = 0.6$: $R^2$,a,b | $\alpha = 0.8$: $R^2$,a,b |
|---|---|---|---|
| $N = 20$, $M = 50$ | 0.9992,1.81,0.52 | 0.9999,2.21,0.70 | 0.9999,2.69,0.70 |
| $N = 100$, $M = 450$ | 0.9998,1.53,1.24 | 0.9997,1.27,1.01 | 0.9994,3.67,1.28 |
| $N = 500$, $M = 35750$ | 0.9999,2.28,1.16 | 0.9994,2.22,1.03 | 0.9986,3.28,1.43 |
| $N = 2000$, $M = 50,000$ | 0.9996,1.88,1.09 | 0.9999,2.43,1.07 | 0.9999,3.57,1.63 |
| $N = 10,000$, $M = 40,000$ | 0.9999,2.61,1.20 | 0.9999,3.14,1.60 | 0.9999,3.60,1.58 |
| $N = 35,000$, $M = 750,000$ | 0.9999,2.32,1.24 | 0.9999,2.61,1.33 | 0.9999,3.24,1.47 |
| $N = 125,000$, $M = 500,000$ | 0.9999,2.56,1.21 | 0.9999,2.73,2.32 | 0.9999,3.39,1.63 |
| $N = 153,036$, $M = 4,152,739$ | 0.9999,2.39,1.14 | 0.9999,2.65,1.41 | 0.9999,3.46,1.75 |

social networks are available for public use[1] as part of studying the spread of Ebola in Africa. The fitting of $R(x)$ is reported in Figures 4-7. The $R^2$ values are close to 1 with minimum value of 0.9984 and maximum value of 0.9999 showing that the error function matches the reliability polynomial well and the parameters $a$ and $b$ are thus suitable for characterizing the reliability polynomial and, by extension, the structure of graphs.

[1] `http://vbi.vt.edu/ndssl/ebola`

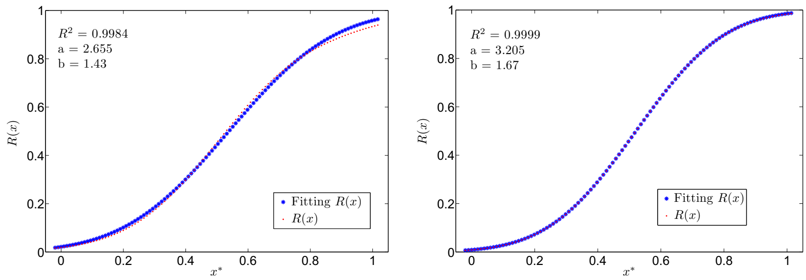


**Fig. 4.** Plot of $R(x)$ and the error function fit with $x^*$ for the NRV social network for reliability rules (a) $ExpX - 0.1$ (left panel), and (b) $ExpX - 0.2$ (right panel)

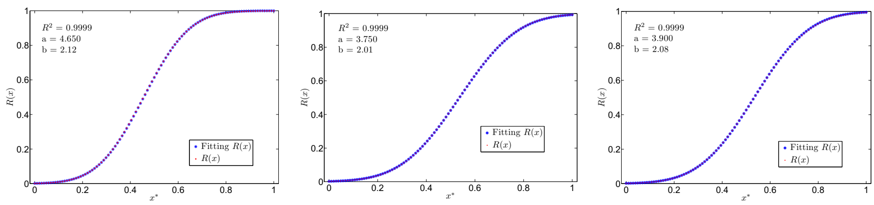


**Fig. 5.** Plot of $R(x)$ and the error function fit with $x^*$ for Mexico City social network for reliability rules (a) $ExpX - 0.05$ (left panel), (b) $ExpX - 0.1$ and (c) $ExpX - 0.2$ (right panel)

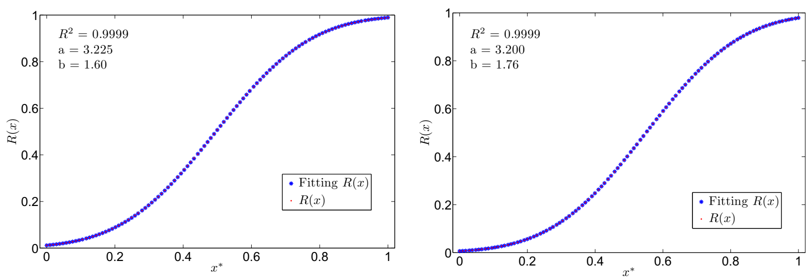


**Fig. 6.** Plot of $R(x)$ and the error function fit with $x^*$ for the Liberia social network for reliability rules (a) $ExpX - 0.05$ (left panel), and (b) $ExpX - 0.1$ (right panel)

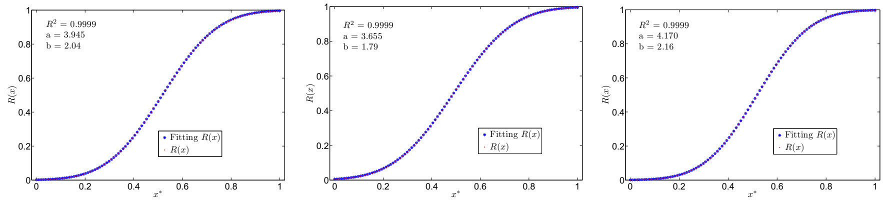


**Fig. 7.** Plot of $R(x)$ and the error function fit with $x^*$ for Sierra Leone social network for reliability rules (a) $ExpX - 0.05$ (left panel), (b) $ExpX - 0.1$ and (c) $ExpX - 0.2$ (right panel)

## 5 Conclusions

We have compared a set of reliability polynomials with both the cumulative distribution function (CDF) of a binomial and the error function. We have reported that the error function yields a better fit to the $P_k$ values than the binomial CDF. We suggest using the parameters of the best-fit error function to characterize $R(x)$. The parameters $a$ and $b$ in the error function change the width and shift the position of the mean of the corresponding Gaussian function, respectively. These values increase with the size of the graph, for several different rules. Finally, we use this method to study the nature of the reliability polynomials of the synthetic social networks for NRV, Mexico City, Sierra Leone and Liberia. We conclude that these two parameters and the values of $k_{min}$ and $k_{max}$, or equivalently or $x_{min}$ and $x_{max}$, for a particular network, summarize the reliability of the network.

We suggest several ways to use the analyses presented here. First, we can use the values $x_{min}$, $x_{max}$, $a$, and $b$ as descriptive statistics for a network that are more informative for many purposes than the usual statistics such as degree distribution, assortativity, etc. Second, they form a set of sufficient statistics for a given feature of diffusive dynamics on a network. Understanding the relationship between network structure and these statistics provides insight into the structure-to-function problem for networks. Finally, this parameterized form for the reliability polynomial can be useful for studying critical point phenomenology in finite-size systems. Here we have described its use for phenomena related to the epidemic transition, but this is just one instance of a percolation transition.

**Acknowledgment.** This work has been partially supported by the Defense Threat Reduction Agency (DTRA) [award number HDTRA1-11-1-0016]; the National Institute of General Medical Sciences of the National Institutes of Health (NIH) [Models of Infectious Disease Agent Study (MIDAS) award number 2U01GM070694-09] and by the National Science Foundation (NSF) [Network Science and Engineering Grant CNS-1011769]. The content is solely the responsibility of the authors and does not necessarily represent the official views of DTRA, the NSF, or the NIH.

## References


1. Moore, E., Shannon, C.: Reliable circuits using less reliable relays. Journal of the Franklin Institute 262, 191–208 (1956)
2. Ahuja, R.K., Magnanti, T.L., Orlin, J.B.: Network flows: theory, algorithms, and applications. Prentice Hall (1993)
3. Ford, L.R., Fulkerson, D.R.: Flows in Networks. Princeton University Press (1962)
4. Colbourn, C.J.: The Combinatorics of Network Reliability. Oxford University Press (1987)
5. Eubank, S., Youssef, M., Khorramzadeh, Y.: Determining and understanding dynamically important differences between complex networks using reliability-induced structural motifs. In: 2013 International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), Complex Networks Workshop, Tokyo, Japan, December 2-5 (2013)
6. Eubank, S., Youssef, M., Khorramzadeh, Y.: Using the network reliability polynomial to characterize and design networks. Journal of Complex Networks, 1–17 (2014)
7. Halloran, M., Vespignani, A., Bharti, N., Feldstein, L., Alexander, K., Ferrari, M., Shaman, J., Drake, J., Porco, T., Eisenberg, J., Valle, S., Lofgren, E., Scarpino, S., Eisenberg, M., Gao, D., Hyman, J., Eubank, S.: Ebola: Mobility data. Science 346 (2014)
8. Youssef, M., Khorramzadeh, Y., Eubank, S.: Network reliability: the effect of local network structure on diffusive processes. Physical Review E 66 (2013)