Relaxation of Gene Network after Injection of Influenza Virus

Shiguo Deng, Henggang Ren, Qianshun Yuan and Huijie Yang
Business School, University of Shanghai for Science and Technology, Shanghai 200093, China.
Email: hjyang@usst.edu.cn

Abstract. The relaxation of gene network after injection of influenza virus is investigated from the perspective of diffusion process. The key idea is to separate each snapshot network into communities, and the identification number of the community in which every node occurs is used to represent the state of the node in this snapshot. From a series of snapshots of network one can construct the trajectories for all the nodes, which form an ensemble. The relaxation is then described by the division and fusion of the communities. It is found that the gene networks for healthy individuals evolves regularly, while that for the individuals affected by influenza virus in a chaotic way.

1. Introduction
The physiological state of an individual evolves in a complicated way that is usually interrupted by disease and the subsequent drug therapy. The gene expressions are sensitive to and used as representative of the physiological state [1, 2]. Recently, an interesting work is conducted to record a series of gene expressions with specific time intervals, which form a multivariate time series [3]. The volunteers are all healthy, whose blood are sampled as reference. Then they are injected with influenza virus, which leads to a division of them into two groups of survivals and sickness. This data stimulate an interesting problem, namely, what are the characteristics of the relaxation of the subjects after the shock of virus injection, and if there exists any early warning signals prior to the occurrence of disease?

In the present paper we map the data to a series of cross-correlation networks of the genes, whose evolutionary behaviours are investigated from the viewpoint of diffusion process. First, we separate the gene expression series into segments and calculate the cross-correlations between the genes respectively. We introduce a threshold to filter out the weak cross-correlations, which results into a series of complex networks to represent the state of the corresponding time durations. Second, the networks are divided into communities. The state of every gene at a specific time duration is described with the identification number of the community it belongs to. From the network series we will obtain trajectories of all the genes, which form an ensemble. By this way the expression series is mapped to a diffusion process.

It is found that the group infected by influenza virus has a complicated diffusion process, namely, the genes in a community in the present step tend to take part in the formations of different communities in the forthcoming step. However, the diffusion process for the survival group tends to behave regular, i.e., the genes clustered in one community tend to occur also in the same community in the next step. This dynamical characteristic provide us clues for prediction.
2. Data and Method

2.1. Gene Expressions for Volunteers Injected with Influenza Virus
The data contains a total of 17 healthy volunteers [3], whose bloods are sampled first as reference. 24 hours later, they are all injected with H3N2 influenza virus, after which for every volunteer in a duration of 108 hours the bloods are sampled 15 times with specified time intervals. The individuals numbered 1, 5, 6, 7, 8, 10, 12, 13 and 15 (totally 9 persons) are infected by the influenza, while the others numbered 2, 3, 4, 9, 11, 14, 16 and 17 survive from infection.

From the gene expression data, we select a total of 416 genes from all of the 11961 genes, which have significant changes in the monitoring duration. For each gene we have an expression series that contains a total of 16 expression levels. We separate the duration into 4 periods, which contains the samples 1-4, 5-8, 9-12, and 13-16, respectively.

2.2. Similarity Networks
In a specific period, the number of expression levels for each gene is 36 in the infected group, and 32 in the survival group, by using of which we can calculate the absolutes value of cross-correlations between all the pairs of genes, respectively. Introducing a threshold to filter out the weak correlations, namely, replacing the correlations larger than it with one and smaller than it zero, one converts the cross-correlations into two similarity networks for the two groups in the corresponding period. To expose the nontrivial structures of the networks, the average degree (the number of neighbours) is selected to be four in all the constructed networks.

The constructed networks are then divided into communities by means of the Newman-Girvan algorithm [4]. A community is denoted with $G(m,t)$ if it is the $m$'th community in the $t$'th period. The trajectory of a specific gene will be $G(a,1),G(b,2),G(c,3)$ and $G(d,4)$, if it occurs in the communities numbered $k=a$, b, c and d in the four periods respectively.

All the trajectories form an ensemble, which form a diffusion process. To find the characteristics of this diffusion process, we define two concepts of the division rate and fusion rate [5], which read,

$$E_{\text{dis}}^t (k,m) \equiv \frac{|G(k,t) \cap G(m,t+1)|}{|G(k,t)|}; \quad E_{\text{fus}}^t (k,m) \equiv \frac{|G(k,t) \cap G(m,t+1)|}{|G(m,t+1)|}$$

They tell us the ratio of genes in the $G(k,t)$ that take part in the formation of $G(m,t+1)$ in the next step, and the ratio of genes in the $G(m,t+1)$ that come from the $G(k,t)$, respectively. If the two ratios are all large, most of the genes in $G(k,t)$ will form the most part of $G(m,t+1)$, i.e., they behave similar and regular. A small value of the measures implies significant division or fusion, called chaotic state.

3. Results
Figure 1 shows the similarity networks for the infected group in the four periods (see the panels (a- d)), and that for the survival group (see the panels (a’-d’)). Figure 2 gives the diffusion processes for the similarity networks. The division rate and fusion rate are marked with grey and blue colours, whose values are proportional to the widths, respectively. For visual convenience, we discard the linkages whose values are less than 0.1.

From figure 2, one can find distinguishable difference between the infected group shown in the panel (a) and the survival group in the panel (b). For example, in the infected group, one can find that $G(2,1)$ splits into several parts that take part mainly in the formations of $G(1,2)$, $G(3,2)$, $G(4,2)$, and $G(8,2)$ in the period 2, most genes from which cluster further to $G(2,3)$ in the next step. Hence, the genes in this community behave similar occasionally in the infected group, i.e., they are in a chaotic state.

Interestingly, all the genes mentioned above cluster also in the $G(1,1)$ in the survival group in panel (b), which conserves in the following periods as $G(1,2)$, $G(1,3)$ and $G(1,4)$.
Figure 1. Similarity networks constructed from gene expressions.

In the panels (a-d) we show the networks in the four periods for the group of survival individuals. In the panels (a’-d’) we show the networks in the four periods for the group of infected individuals. The networks are divided into communities by means of the Newman-Girvan method. Each community is denoted with $G(m,t)$, where $m$ is the identification number of the community and $t$ the period.

4. Conclusions and Discussions
The gene network is used to represent the state of physiology. The division and fusion of communities are used to describe the dynamical process from the perspective of diffusion. It is found that the gene network for individuals affected by influenza virus behaves chaotic, while that for survival individuals regular. And some specific genes are found that display significant behaviours in the two groups. We hope these findings are helpful in biomarkers and early warning signals of disease. The method can be extended straightforwardly to investigate other complex systems such as stock markets, human organisms, social groups and brains.

5. Acknowledgements
The authors thank the supports from the National Natural Science Foundation of China under grant nos. 11875042, 11505114, 11805128 and 10975099, and the Shanghai project for construction of top disciplines under Grant No. USST-SYS-01.
(a) The infected group. (b) The survival group.

Figure 2. The division and fusion of the gene networks.

The division and fusion rates are represented with grey and blue directional linkages. The width of a linkage is proportional to the value of ratio. The 100% near the communities in (b) implies that the genes in $G(1,1)$ form the groups of $G(1,2)$, $G(1,3)$ and $G(1,4)$ in the following periods. In the panels the labels time1, time2, time3, and time4 refer to the periods from 1 to 4.

6. References
[1] Liu R, Wang X, Aihara K, et al. 2014 Med. Res. Rev. 34(3):455
[2] Zhou L, Qiu L, Gu C G, Yang H J 2018 EPL 121:48002
[3] Huang Y, Zaas A K, Rao A 2011 PLoS Genetics 7(8):e1002234
[4] Girvan M, Newman M E J 2006 PNAS 99(12):7821
[5] Deng S G, Qiu L, Yang Y and Yang H J 2016 Physica A 441:62