

Article

## A statistical simulation method for causality inference of Boolean variables

**WenJun Zhang**

School of Life Sciences, Sun Yat-sen University, Guangzhou 510275, China

E-mail: zhwj@mail.sysu.edu.cn, wjzhang@iaees.org

Received 6 August 2020; Accepted 28 September 2021; Published online 6 October 2021; Published 1 December 2021



### Abstract

In present study, a statistical simulation method for causality inference of Boolean variables was proposed. First, I used statistical simulation to generate artificial data of two Boolean variables with known independent and dependent variables. A law was drawn from the simulation analysis of the artificial data. For a set of data of two Boolean variables, a randomization method was proposed and used to test the statistical significance of the Boolean correlation measure (point correlation, quartile correlation, or Jaccard correlation, etc.). The causality inference was then conducted to observed data based on the law. Finally, the statistical simulation was used to determine the statistic significance of the causality. Full Matlab codes were presented also.

**Keywords** causality inference; correlation; Boolean variables; randomization; statistical simulation; non-parametric statistics.

**Network Biology**  
ISSN 2220-8879  
URL: <http://www.iaees.org/publications/journals/nb/online-version.asp>  
RSS: <http://www.iaees.org/publications/journals/nb/rss.xml>  
E-mail: [networkbiology@iaees.org](mailto:networkbiology@iaees.org)  
Editor-in-Chief: WenJun Zhang  
Publisher: International Academy of Ecology and Environmental Sciences

### 1 Introduction

Causality inference is a frontier area in science. So far few successful theories have been made in causality inference of variables. Causality will result in correlation between two variables. We can conduct causality infernce only the two variables are correlated with each other. Theories and applications of correlations have been well studied, including that in network biology (Qi and Zhang, 2003; Kuang and Zhang, 2011; Huang and Zhang, 2012; Jiang and Zhang, 2015a, b; Zhang, 2007, 2011b, 2012a, 2014-2018, 2021; Zhang and Zhang, 2019; Xin and Zhang, 2020). There are many correlation measures, among which Pearson correlation, Spearman correlation, etc., are for interval variables, and point correlation, Jaccard coefficient, etc., are for Boolean (binary) variables. In addition to the methods of parametrical statistics, statistical simulation (randomization) methods are widely used to make statistical inferences (Solow, 1993; Manly, 1997; Zhang and Schoenly, 1999; Zhang, 2010, 2011a). In present study, a statistical simulation (randomization) method for causality inference of Boolean variables was proposed. Full Matlab codes were presented also.

## 2 Correlations and Statistic Tests

Causality will result in correlation between two variables. We can conduct causality inference only the two variables are correlated with each other. There are many correlation measures, among which Pearson correlation, Spearman correlation, etc., are for interval variables, and point correlation, Jaccard coefficient, etc., are for Boolean (binary) variables (Qi and Zhang, 2003; Zhang, 2007, 2012a, 2014, 2015, 2016, 2017, 2018, 2021).

### 2.1 Correlation measures of Boolean variables

There are a lot of correlation measures for Boolean variables. Here I use five representative ones as follows.

#### 2.1.1 Point correlation

Suppose there are  $n$  pairs of Boolean data. The point correlation between Boolean variables  $x$  and  $y$  is (Qi and Zhang, 2003; Zhang, 2007, 2012a, 2015, 2016, 2017, 2018, 2021; Zhang et al., 2014)

$$r=(ad-bc)/((a+b)(c+d)(a+c)(b+d))^{1/2}$$

where  $-1 \leq r \leq 1$ , both variable  $x$  and variable  $y$  take values 0 or 1.  $a$  is number of both variable  $x$  and variable  $y$  take value 0,  $b$  is number of variable  $x$  takes 0 and variable  $y$  takes 1,  $c$  is number of variable  $x$  takes 1 and variable  $y$  takes 0, and  $d$  is number of both variable  $x$  and variable  $y$  take value 1.  $n=a+b+c+d$ . A greater and positive  $r$  means a positive point correlation between variables  $x$  and  $y$ ; a greater and negative  $r$  means a negative point correlation between variables  $x$  and  $y$ .

#### 2.1.2 Quartile correlation

Quartile correlation is defined by (Zhang, 2016, 2018):

$$r=\sin((a+d-(b+c))/(a+b+c+d)*\pi/2)$$

where  $-1 \leq r \leq 1$ ;  $a$ ,  $b$ ,  $c$  and  $d$  follow the definitions above.

#### 2.1.3 Cosine correlation

Cosine correlation is defined by (Zhang, 2016, 2018):

$$r=(a^2*d^2/((a+b)(a+c)(b+d)(c+d)))^{1/2}$$

where  $0 \leq r \leq 1$ ;  $a$ ,  $b$ ,  $c$  and  $d$  follow the definitions above.

#### 2.1.4 Cosine+ correlation

$$r=(a^2/((a+b)(a+c)))^{1/2}$$

where  $0 \leq r \leq 1$ ;  $a$ ,  $b$ , and  $c$  follow the definitions above.

#### 2.1.5 Jaccard correlation

Jaccard correlation is defined by (Qi and Zhang, 2003; Zhang, 2007, 2012a, 2014, 2015, 2016, 2017, 2018, 2021):

$$r=(d-(c+b))/(d+c+b)$$

where  $-1 \leq r \leq 1$ ;  $b$ ,  $c$  and  $d$  follow the definitions above.

## 2.2 Detection of correlation between two Boolean variables

Following the principle of randomization methods (Manly, 1997; Solow, 1993; Zhang, 2007, 2010, 2011a; Zhang and Schoenly, 1999), here I propose a method to test the statistical significance of the Boolean correlation measures as described above. Suppose that there are  $n$  pairs of observed data for two Boolean variables  $x$  and  $y$ , as demonstrated below:

$$\begin{array}{l} x \quad 0 \ 1 \ 1 \ 0 \ 0 \ 0 \ 1 \ 0 \ 0 \ 1 \ 1 \ 0 \ \cdots \\ y \quad 1 \ 1 \ 0 \ 1 \ 0 \ 0 \ 0 \ 1 \ 0 \ 0 \ 1 \ 0 \ \cdots \end{array}$$

We assume that there are  $n_x$  and  $n_y$  of 1 in vector  $x$  and vector  $y$  respectively. First, calculate the practical correlation  $r$  of  $x$  and  $y$ . Second, randomly re-assign  $n_x$  and  $n_y$  of 1 in  $x$  and  $y$  respectively, and calculate the theoretical correlation  $r$  of re-arranged  $x$  and  $y$ . Repeat the procedure  $s$  times, and find the times of absolute theoretical correlations greater than the absolute practical correlation,  $w$ . If  $w/s < p$ , here  $p=0.01, 0.001$ , etc., then we conclude that the practical correlation  $r$  between  $x$  and  $y$  is statistically significant at the statistical level  $p$ .

## 3 Causality Inference of Two Boolean Variables

There is a possible causality between two Boolean variables if their correlation is statistically significant. To exploit the general law of causality and correlation, we need to construct the artificial data of two Boolean variables, known the independent variable,  $x$ , and dependent variable,  $y$ .

### 3.1 Causality principle of Boolean variables

Suppose that the significant correlation between two Boolean variables  $x$  and  $y$  is confirmed. We assume that there is likely a causality between two Boolean variables  $x$  and  $y$ , and furthermore,  $x$  is the independent variable and  $y$  is the dependent variable. In a set of sample data, as demonstrated above,  $x=1$  will most likely result in  $y=1$  at a greater probability. And,  $y=1$  that have not been resulted from  $x=1$  occurs occasionally at a smaller probability and meet  $x=0$  or  $x=1$ . The latter stochastic  $y=1$  is produced from the stochasticity of variable  $y$  and the dependence of  $y$  on other unknown variables.

### 3.2 Relationship between causality and statistic parameters

#### 3.2.1 Statistical simulation

From the Boolean correlation measures above, it is apparent that the only valuable parameters for causality inference are  $b$  and  $c$ . The parameters,  $a$  and  $d$ , are not able to provide any valuable information on the inference of between-variable causality. I try to find the deterministic relationship between ( $b, c$ ) and causality by statistical simulation.

To avoid infinity (Inf, value/0) or non-value (NaN, 0/0), only quartile correlation is used in the simulation. In a simulation, construct the data of independent variable  $x$  and dependent variable  $y$  following the method above. Second, calculate the quartile correlation  $r$ , and  $b, c$ . Repeat the procedure many times, each time construct the new data of independent variable  $x$  and dependent variable  $y$  with random data sizes and dependency probabilities and dependency errors, etc., and record  $r$ , and  $b, c$ . Finally, calculate the mean  $r$ , mean  $b$  and mean  $c$ , and test the statistic difference between mean  $b$  and mean  $c$ . Meanwhile, calculate the Pearson correlation between absolute  $r$  and  $c > b$  ( $c > b$ : 1;  $c \leq b$ : 0) and the Pearson correlation between absolute  $r$  and  $c - b$ , and make statistic tests on the Pearson correlations.

The full Matlab codes of the statistical simulation, xyGen, for finding relationship between causality and statistic parameters are as follows (see supplementary material also):

```

clear
sel=1;      %sel=1: for constructing positive dependency; sel=2: for constructing negative dependency
sig=0.01;   %Statistic significance level for detecting Pearson correlation between Boolean correlation r and c-b
xdevi=0.2;  %Deviation of dependency of y on x, e.g., 0.1, 0.2, 0.3, etc.
yerr=0.7;   %1-yerr: stochastic error of y, determined by random error and the influence of other unknown variables,
%e.g., 0.6, 0.7, 0.8, etc.
m=200;      %For determining the maximum size of Boolean variables x and y
sim=20000;  %Number of simulations (randomizations); generally, sim>=m*100
res=zeros(sim,7);
nn=zeros(sim,1);
for i=1:sim
n=floor(m*rand()+10); %The size, n, can be fixed, e.g., n=m
nn(i)=n;
xprob=rand()-xdevi;
x=zeros(n,1);
y=zeros(n,1);
xbas=rand(n,1);
x=xbas>xprob;
if (sel==1) yx=xbas>(xprob+xdevi);
elseif (sel==2) yx=xbas<(xprob-xdevi); end;
yran=rand(n,1)>yerr;
y=yx | yran;
[r,a,b,c,d]=quartilecorr(x,y);
res(i,:)=[sum(x) sum(y) r a b c d];
end
means_of_b_and_c=mean(res(:,5:6))
[h,p_value_for_difference_between_b_and_c,v]=ttest2(res(:,5),res(:,6),sig);
p_value_for_difference_between_b_and_c
if (h==0)
sprintf(['The difference between b and c is not statistically significant at p=',num2str(sig),'\n'])
else sprintf(['The difference between b and c is statistically significant (p=', num2str(p_value_for_difference_between_b_and_c),
')\n'])
end
r_mean=mean(res(i,3))
Pearson_correlation_between_absolute_r_and_c_greater_than_b=corr(abs(res(:,3)),double(res(:,6)>res(:,5)))
tvalue=abs(Pearson_correlation_between_absolute_r_and_c_greater_than_b)/sqrt((1-Pearson_correlation_between_absolute_r_and_c_greater_than_b.^2)/(n-2));
alpha=(1-tcdf(tvalue,mean(nn)-2))*2;
if (alpha<=sig)
sprintf(['Pearson correlation between absolute r and c>b is statistically significant (p=',num2str(alpha),')\n'])
else sprintf(['Pearson correlation between absolute r and c>b is not statistically significant at p=',num2str(sig),'\n'])
end
Pearson_correlation_between_absolute_r_and_c_minus_b=corr(abs(res(:,3)),res(:,6)-res(:,5))
tvalue=abs(Pearson_correlation_between_absolute_r_and_c_minus_b)/sqrt((1-Pearson_correlation_between_absolute_r_and_c_minus_b.^2)/(n-2));

```

```

alpha=(1-tcdf(tvalue,mean(nn)-2))*2;
if (alpha<=sig)
sprintf(['Pearson correlation between absolute r and c-b is statistically significant (p=',num2str(alpha),')\n'])
else sprintf(['Pearson correlation between absolute r and c-b is not statistically significant at p=',num2str(sig),'\n'])
end

function [r,a,b,c,d]=quartilecorr(x,y) %x and y: two column vectors to be tested.
if (max(size(x))~=max(size(y)))
error('Array sizes do not match. ');
end
if ((min(size(x))~=1) | (min(size(y))~=1))
error('Both x and y are vectors');
end
a=sum((x==0) & (y==0));
b=sum((x==0) & (y~=0));
c=sum((x~=0) & (y==0));
d=sum((x~=0) & (y~=0));
r=sin((a+d-(b+c))/(a+b+c+d)*pi/2);

```

### 3.2.2 Law found from statistical simulation

The results of statistical simulation indicated that for positive causality ( $r > 0$ ;  $x$ : independent variable,  $y$ : dependent variable), there is a significant positive Pearson correlation between absolute  $r$  and  $c > b$  ( $c > b$ : 1;  $c \leq b$ : 0), and between absolute  $r$  and  $c < b$ ; for negative causality ( $r < 0$ ;  $x$ : independent variable,  $y$ : dependent variable), there is a significant positive Pearson correlation between absolute  $r$  and  $c \leq b$  ( $c > b$ : 0;  $c \leq b$ : 1), and between absolute  $r$  and  $b < c$ .

### 3.3 Statistical simulation for causality inference based on observed Boolean data

According to the law above, a significant  $r$  and  $c > b$  ( $r > 0$ ;  $x$ : independent variable,  $y$ : dependent variable) or  $c < b$  ( $r > 0$ ;  $y$ : independent variable,  $x$ : dependent variable) can be used as the criteria for possible causality of Boolean variables.

First, assume that the Boolean correlation between the two Boolean variables is statistically significant. Following the principle of randomization tests (Manly, 1997; Solow, 1993; Zhang, 2007, 2010, 2011a; Zhang and Schoenly, 1999), I try to use statistical simulation to find the causality based on observed Boolean data. In each randomization process, two re-arranged Boolean variables,  $xx$  and  $yy$ , are generated based on the observed data,  $x$  and  $y$ . The elements with shared 0 and 1 in original observed data are reserved in re-arranged  $x$  and  $y$ . And other elements are proportionally randomized re-arranged as follows:

$$v_1 = c / (b + c)$$

$$v_2 = b / (b + c)$$

$$z = \text{rand}(n)$$

$$xy = x \vee y$$

$$xx = (x \wedge y) \vee (xy \wedge (z > 1 - v_1))$$

$$yy = (x \wedge y) \vee (xy \wedge (z \leq v_2))$$

Then calculate and record  $b$  and  $c$ . Repeat the procedure many times, finally calculate the proportions (i.e., calculated  $p$  values,  $p_1, p_2$ ) of  $c > b$  and  $c \leq b$ .

For  $r > 0$ , if  $c > b$  and  $p_2 < p$ , where  $p$  is the given significance level, then  $x$  is the independent variable and  $y$  is the dependent variable; or else if  $c < b$  and  $p_1 < p$ , then  $y$  is the independent variable and  $x$  is the dependent variable. For  $r < 0$ , if  $c < b$  and  $p_1 < p$ , then  $x$  is the independent variable and  $y$  is the dependent variable; or else if  $c > b$  and  $p_2 < p$ , then  $y$  is the independent variable and  $x$  is the dependent variable.

The full Matlab codes, `causalInferBool`, of the statistical simulation for causality inference based on observed Boolean data are as follows (see supplementary material also):

```
clear
xyd=input('Input the Excel file name of raw data (e.g., xyd.xls: xyd=(dij)n×2, dij=0 or 1, i=1,2,...,n; j=1,2. In the file, column 1
is for variable 1 and column 2 is for variable 2): ','s');
sel=input('Choose a correlation measure (1: point correlation; 2: quartile correlation; 3: cosine correlation; 4: cosine+ correlation;
5: Jaccard correlation): ');
p=input('Input the statistical significance level p for correlation inference (e.g., 0.001): ');
sig=input('Input the statistical significance level p for causality inference (e.g., 0.05, 0.1, 0.2, etc.): ');
sim=input('Input the number of simulations (e.g., 10000): ');
xyd=xlsread(xyd);
n=size(xyd,1);
x=xyd(:,1);
y=xyd(:,2);
switch (sel)
    case 1
        [r,a,b,c,d]=pointcorr(x,y); str='point correlation';
    case 2
        [r,a,b,c,d]=quartilecorr(x,y); str='quartile correlation';
    case 3
        [r,a,b,c,d]=cosinecorr(x,y); str='cosine correlation';
    case 4
        [r,a,b,c]=cosinepluscorr(x,y); str='cosine+ correlation';
    case 5
        [r,b,c,d]=jaccard(x,y); str='Jaccard correlation';
end
if (isnan(r) | (r==Inf) | (r==-Inf))
    sprintf(['The ',str,' measure is not valid. Try to use another correlation measure.\n'])
    pause();
end
nx=sum(x);
ny=sum(y);
ww=0;
ss=0;
for s=1:sim
    xx=zeros(n,1);
    yy=zeros(n,1);
    idx=randperm(n);
```

```

idy=randperm(n);
for i=1:nx
xx(idx(i))=1;
end
for i=1:ny
yy(idy(i))=1;
end
switch (sel)
    case 1
        [rs,as,bs,cs,ds]=pointcorr(xx,yy);
    case 2
        [rs,as,bs,cs,ds]=quartilecorr(xx,yy);
    case 3
        [rs,as,bs,cs,ds]=cosinecorr(xx,yy);
    case 4
        [rs,as,bs,cs]=cosinepluscorr(xx,yy);
    case 5
        [rs,bs,cs,ds]=jaccard(xx,yy);
end
if (isnan(rs) | (rs==Inf) | (rs==-Inf)) continue; end
ss=ss+1;
if (((r>0) & (rs>r)) | ((r<0) & (rs<r))) ww=ww+1; end
end
id=0;
if ((ww/ss)<p)
sprintf(['There is a significant ',str,' (r=',num2str(r),') between two variables (p=',num2str(ww/ss),')\n'])
id=1;
else sprintf(['There is not significant ',str,' (r=',num2str(r),') between two variables (p=',num2str(ww/ss),')\n'])
sprintf(['So, causality may not exist between two variables based on ',str,'\n'])
end
v1=c/(b+c);
v2=b/(b+c);
for s=1:sim
z=rand(n,1);
xy=x | y;
xx=(x & y) | (xy & (z>1-v1));
yy=(x & y) | (xy & (z<=v2));
bs=sum((xx==0) & (yy~=0));
cs=sum((xx~=0) & (yy==0));
bc(s,:)= [bs cs];
end
p1=sum(bc(:,2)>bc(:,1))/sim;
p2=sum(bc(:,2)<=bc(:,1))/sim;
if (r>0)
if ((c>b) & (id==1) & (p2<sig))

```

```

sprintf(['The variable for column 1 is the independent variable, and the variable for column 2 is the dependent variable
(p=',num2str(p2),')\n'])
elseif ((c<b) & (id==1) & (p1<sig))
sprintf(['The variable for column 2 is the independent variable, and the variable for column 1 is the dependent variable
(p=',num2str(p1),')\n'])
else sprintf(['However, causality may not exist between two variables (p=',num2str(sig),')\n'])
end
elseif (r<0)
if ((c<b) & (id==1) & (p1<sig))
sprintf(['The variable for column 1 is the independent variable, and the variable for column 2 is the dependent variable
(p=',num2str(p1),')\n'])
elseif ((c>b) & (id==1) & (p2<sig))
sprintf(['The variable for column 2 is the independent variable, and the variable for column 1 is the dependent variable
(p=',num2str(p2),')\n'])
else sprintf(['However, causality may not exist between two variables (p=',num2str(sig),')\n'])
end
end

```

```

function [r,a,b,c,d]=pointcorr(x,y)
%x and y: two column vectors to be tested.
if (max(size(x))~=max(size(y)))
error('Array sizes do not match.');
```

```

end
if ((min(size(x))~=1) | (min(size(y))~=1))
error('Both x and y are vectors');
```

```

end
a=sum((x==0) & (y==0));
b=sum((x==0) & (y~=0));
c=sum((x~=0) & (y==0));
d=sum((x~=0) & (y~=0));
r=(a*d-b*c)/sqrt((a+b)*(c+d)*(a+c)*(b+d));

```



```

function [r,a,b,c,d]=cosinecorr(x,y)           %x and y: two column vectors to be tested.
if (max(size(x))~=max(size(y)))
    error('Array sizes do not match.');
```

end

```

if ((min(size(x))~=1) | (min(size(y))~=1))
    error('Both x and y are vectors');
```

end

```

a=sum((x==0) & (y==0));
b=sum((x==0) & (y~=0));
c=sum((x~=0) & (y==0));
d=sum((x~=0) & (y~=0));
r=sqrt(a^2*d^2/((a+b)*(a+c)*(b+d)*(c+d)));
```

```

function [r,a,b,c]=cosinepluscorr(x,y)       %x and y: two column vectors to be tested.
if (max(size(x))~=max(size(y)))
    error('Array sizes do not match.');
```

end

```

if ((min(size(x))~=1) | (min(size(y))~=1))
    error('Both x and y are vectors');
```

end

```

a=sum((x==0) & (y==0));
b=sum((x==0) & (y~=0));
c=sum((x~=0) & (y==0));
r=sqrt(a^2/((a+b)*(a+c)));
```

```

function [r,b,c,d]=jaccard(x,y)
%x and y: two column vectors to be tested.
if (max(size(x))~=max(size(y)))
    error('Array sizes do not match.');
```

end

```

if ((min(size(x))~=1) | (min(size(y))~=1))
    error('Both x and y are vectors');
```

end

```

b=sum((x==0) & (y~=0));
c=sum((x~=0) & (y==0));
d=sum((x~=0) & (y~=0));
r=(d-(c+b))/(d+c+b);
```

A set of theoretical data were used to validate the method and the goodness of the method was confirmed.

#### 4 Discussion

To increase the reliability of causality inference, the size of Boolean data should be large enough. For a large data, the present method can be used on randomly segmented data blocks (i.e., bootstrap method) in order to draw more reliable conclusion from multiple results. The present method has been tested and validated with

theoretical data. However, future applications and further validations are expected and possible improvements on the method may be expected.

### Acknowledgment

I am thankful to the support of The National Key Research and Development Program of China (2017YFD0201204), from China, and Discovery and Crucial Node Analysis of Important Biological and Social Networks (2015.6-2020.6), from Yangling Institute of Modern Agricultural Standardization, China.

### References

- Huang JQ, Zhang WJ. 2012. Analysis on degree distribution of tumor signaling networks. *Network Biology*, 2(3): 95-109
- Jiang LQ, Zhang WJ. 2015a. Determination of keystone species in CSM food web: A topological analysis of network structure. *Network Biology*, 5(1): 13-33
- Jiang LQ, Zhang WJ. 2015b. Effects of parasitism on robustness of food webs. *Selforganizology*, 2(2): 21-34
- Kuang WP, Zhang WJ. 2011. Some effects of parasitism on food web structure: a topological analysis. *Network Biology*, 1(3-4): 171-185
- Manly BFJ. 1997. *Randomization, Bootstrap and Monte Carlo Methods in Biology* (2nd Edition). Chapman & Hall, London, UK
- Qi YH, Zhang WJ. 2003. CorreDetector: A network sharing software used in correlation analysis of information. *Journal of Information*, 22(Suppl.): 266-268
- Solow AR. 1993. A simple test for change in community structure. *Journal of Animal Ecology*, 62: 191-193
- Xin SH, Zhang WJ. 2020. Construction and analysis of the protein-protein interaction network for the olfactory system of the silkworm *Bombyx mori*. *Archives of Insect Biochemistry and Physiology*, 105(3): e21737
- Zhang GL, Zhang WJ. 2019. Protein-protein interaction network analysis of insecticide resistance molecular mechanism in *Drosophila melanogaster*. *Archives of Insect Biochemistry and Physiology*, 100(1): e21523
- Zhang WJ. 2007. Computer inference of network of ecological interactions from sampling data. *Environmental Monitoring and Assessment*, 124: 253-261
- Zhang WJ. 2010. *Computational Ecology: Artificial Neural Networks and Their Applications*. World Scientific, Singapore
- Zhang WJ. 2011a. A Java program to test homogeneity of samples and examine sampling completeness. *Network Biology*, 1(2): 127-129
- Zhang WJ. 2011b. Constructing ecological interaction networks by correlation analysis: hints from community sampling. *Network Biology*, 1(2): 81-98
- Zhang WJ. 2012a. *Computational Ecology: Graphs, Networks and Agent-based Modeling*. World Scientific, Singapore
- Zhang WJ. 2012b. How to construct the statistic network? An association network of herbaceous plants constructed from field sampling. *Network Biology*, 2(2): 57-68
- Zhang WJ. 2014. Interspecific associations and community structure: A local survey and analysis in a grass community. *Selforganizology*, 1(2): 89-129
- Zhang WJ. 2015. Calculation and statistic test of partial correlation of general correlation measures. *Selforganizology*, 2(4): 65-77
- Zhang WJ. 2016. *Selforganizology: The Science of Self-Organization*. World Scientific, Singapore

- Zhang WJ. 2017. Network pharmacology of medicinal variables and functions of Chinese herbal medicines: (II) Relational networks and pharmacological mechanisms of medicinal variables and functions of Chinese herbal medicines. *Network Pharmacology*, 2(2): 38-66
- Zhang WJ. 2018. *Fundamentals of Network Biology*. World Scientific Europe, London, UK
- Zhang WJ. 2021. A web tool for generating user-interface interactive networks. *Network Biology*, 11(4): 247-262
- Zhang WJ, Schoenly KG. 1999. IRRI Biodiversity Software Series. III. BOUNDARY: A Program for Detecting Boundaries in Ecological Landscapes. IRRI Technical Bulletin No.3. International Rice Research Institute, Manila, Philippines
- Zhang WJ, Wang R, Zhang DL, et al. 2014. Interspecific associations of weed species around rice fields in Pearl River Delta, China: A regional survey. *Selforganizology*, 1(3-4): 143-205