Article

Causality inference of nominal variables: A statistical simulation method

WenJun Zhang
School of Life Sciences, Sun Yat-sen University, Guangzhou 510275, China
E-mail: zhwj@mail.sysu.edu.cn, wjzhang@iaees.org

Received 1 October 2021; *Accepted 5 October 2021*; *Published online 13 October 2021; Published 1 December 2021*
(cc) BY

Abstract
In present study I proposed a statistical simulation method for causality inference of nominal variables (i.e., categorical variables). A new correlation measure for nominal variables, association coefficient, is firstly proposed also. A statistical simulation method was developed to generate artificial data of nominal variables with known causality. The law was then drawn from the simulation analysis of the artificial data. For a set of data of two nominal variables, the randomization method was first used to test the statistical significance of the nominal correlation measure, and then the statistical simulation was used to determine the causality and its statistic significance of two nominal variables. Full Matlab codes of the method were presented.

Keywords causality; inference; correlation; nominal variables; contingency measures; association coefficient; randomization; statistical simulation; non-parametric statistics.

Computational Ecology and Software
ISSN 2220-721X
URL: http://www.iaees.org/publications/journals/ces/online-version.asp
RSS: http://www.iaees.org/publications/journals/ces/rss.xml
E-mail: ces@iaees.org
Editor-in-Chief: WenJun Zhang
Publisher: International Academy of Ecology and Environmental Sciences

1 Introduction

Causality inference between two variables is a hot topic in science. It is well known that causality will result in correlation between two variables. Causality inferernce two variables can be conducted only they correlated with each other. Theories and applications of correlations have been well studied, including those in biology and ecology (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, and contingency coefficients, are for nominal (categorical) variables. In addition to various methods in parametrical statistics, statisticall simulation methods are widely used to make statistical inferences (Solow, 1993; Manly, 1997; Zhang and Schoenly, 1999; Zhang, 2010, 2011a).

So far there are seldom successful parametrical statistic methods for causality inference of variables. In terms of Boolean variables, I have developed a system of statistical simulation methods to make causality inference (Zhang, 2021). However, the causality inference of nominal variables is expected to be much complicated compared that for Boolean variables. In present study, I proposed a statistical simulation method for causality inference of nominal variables. Full Matlab codes of the method were presented for practical uses.

2 Correlations and Statistic Tests of Nominal Variables

2.1 Correlation measures of nominal variables

As mentioned earlier (Zhang, 2021), causality will result in correlation between two variables. Causality inferernce is available only the two variables are correlated with each other. There are a lot of correlation measures for nominal variables *x* and *y*. In present study, I use five nominal correlation (contingency) measures as follows (Zhang and Fang, 1982), in which the association coefficient is firstly proposed here:

(1) Association coefficient

$r=S^2/(S^2+\bar{X})$

(2) Contingency coefficient

$r=(w^2/(w^2+n))^{1/2}$

(3) Contingency coefficient I

$r=(w^2/(n*((p-1)(q-1))^{1/2}))^{1/2}$

(4) Contingency coefficient II

$r=(w^2/(n*max(p-1,q-1)))^{1/2}$

(5) Contingency coefficient III

$r=(w^2/(n*min(p-1,q-1)))^{1/2}$

where $0\leq r\leq 1$; nominal variable *x* has *p* types of unique qualitative values, u_1, u_2, …,u_p, and nominal variable *y* has *q* types of unique qualitative values, v_1, v_2,…, v_q; n_{kl} is the number of element pairs with variable $x=u_k$ and variable $y=v_l$, k=1,2,...,p; l=1,2,...,q, and

$\bar{X}=\sum_{i=1}^{p}\sum_{j=1}^{q}n_{ij}/(p*q)$

$S^2=\sum_{i=1}^{p}\sum_{j=1}^{q}(n_{ij}-\bar{X})^2/(p*q)$

$w^2=n(\sum_{i=1}^{p}\sum_{j=1}^{q}n_{ij}^2/(n_in_j)-1)$

$n=\sum_{i=1}^{p}n_i$

$n_i=\sum_{j=1}^{q}n_{ij}$

$n_j=\sum_{i=1}^{p}n_{ij}$

2.2 Detection of correlation between two nominal variables

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

x	4	2	2	4	2	3	3	4	2	1	…
y	3	1	1	3	4	2	2	4	1	2	…

where variable x has p types of unique qualitative values variable y has q types of unique qualitative values, and it is better to meet $m>>max(p, q)$. First, calculate the practical correlation r of x and y. Second, randomly re-assign elements 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 theoretical correlations greating than the practical correlation, w. If $w/s<\alpha$, here α=0.01, 0.001, etc., then we conclude that the practical correlation r between variables x and y is ststistically significant at the statistical level α.

3 Causality Inference of Two Nominal variables

Causality between two nominal variables can be expected if their correlation is statistically significant. In order to find the general law of causality and correlation between nominal variables, here I will construct the artificial data of two nominal variables, from the independent variable x, to dependent variable y.

3.1 Causality principle of nominal variables

We suppose that the significant correlation between two nominal variables, x and y, has been confirmed. Further, assume that the causality exists between two nominal variables, x and y, and x is the independent variable and y is the dependent variable. Set a function mapping from p types of unique values, u_1, u_2, …,u_p, of variable x, to q types of unique values, v_1, v_2,…, v_q, of of variable y. For a value in independent variable x, the value in dependent variable y will most likely be the mapped value from x, mapped with a greater probability. The random components in the mapping rule represent the stochastic errors in variable y following variable x.

3.2 Relationship between causality and statistic parameters

3.2.1 Statistical simulation

The deterministic relationship between causality and statistic parameters can be exploited by using statistical simulation.

In a statistical simulation, first construct the data of independent variable x and dependent variable y following the method above. Second, calculate the correlation r, and record r, mean of the sum of columns (y) of n_{ij} (V_y), mean of the sum of rows (x) of n_{ij} (V_x), and the ratio of variance of the sum of columns (y) of n_{ij} *vs*. variance of the sum of rows (x) of n_{ij} (V_{yx}). Repeat the procedure many times, each time we construct the new data of independent variable x and dependent variable y, with random data sizes, p and q and the corresponding unique values, and dependency probalbilities, where

$$V_y=\sum_{j=1}^q\sum_{i=1}^p n_{ij}/q$$
$$V_x=\sum_{i=1}^p\sum_{j=1}^q n_{ij}/p$$
$$V_{yx}=(\sum_{j=1}^q(\sum_{i=1}^p n_{ij}-V_y)^2/q)/(\sum_{i=1}^p(\sum_{j=1}^q n_{ij}-V_x)^2/p)$$

Finally, calculate the mean r, the proportion of V_{yx} being less than 1 (i.e., $P(V_{yx}<1)$), the Pearson correlation between r and $V_y<V_x$, and the Pearson correlation between r and V_x-V_y, and make statistic tests on the two Pearson correlations.

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

```
clear;
sel=1;
sig=0.001;      %Statistic significance level for detecting Pearson correlation between nominal variables' correlation r and some statistic measure
yprob=0.5;      %Basic probabiliy of y following x in a given pattern
```

```
m=200;          %For determining the maximum size of nominal variables x and y
sim=5000;       %Number of simulations (randomizations)
for s=1:sim
mn=floor(m*rand()+100);    %The size, mn, can be fixed, e.g., mn=m
x=zeros(mn,1);
y=zeros(mn,1);
p=floor(rand()*mn/20+3);
q=floor(rand()*mn/20+3);
pc=randperm(p);
qc=randperm(q);
if (p>q)
for j=1:p-q
b(j)=floor(rand()*q+1);
end
qc=[qc b(1:p-q)];
elseif (q>p)
for j=1:q-p
b(j)=floor(rand()*p+1);
end
pc=[pc b(1:q-p)];
end
nn=size(pc,2);
x=floor(rand(mn,1)*p+1);
yp=rand(mn,1)+yprob;
for j=1:mn
temp=rand();
for k=1:nn
if ((x(j)==pc(k)) & (temp<yp(j))) y(j)=qc(k); break;
elseif ((x(j)==pc(k)) & (temp>=yp(j))) y(j)=qc(floor(rand()*nn+1)); break;
end; end; end
[rs,nij,nx,ny,str]=nominalcorr(sel,x,y);
vy=mean(sum(nij));
vx=mean(sum(nij'));
vyx=var(sum(nij))/var(sum(nij'));
res(s,:)=[rs vy vx vyx];
end
varyVSvarx=sum(res(:,4)<1)/sim                          %Prob of Vyx<1
r=mean(res(:,1))
Pearson_corr_lt=corr(res(:,1),res(:,2)<res(:,3))              %r and Vy<Vx
tvalue=abs(Pearson_corr_lt)./sqrt((1-Pearson_corr_lt.^2)/(sim-2));
alpha=(1-tcdf(tvalue,mean(sim)-2))*2;
if (alpha<=sig)
sprintf(['Pearson correlation between r and vy<vx is statistically significant (p=',num2str(alpha),')\n'])
else sprintf(['Pearson correlation between r and vy<vx is not statistically significant at p=',num2str(sig),'\n'])
end
```

```
Pearson_corr_minus=corr(res(:,1),res(:,3)-res(:,2))             %r and Vx-Vy
tvalue=abs(Pearson_corr_minus)./sqrt((1-Pearson_corr_minus.^2)/(sim-2));
alpha=(1-tcdf(tvalue,mean(sim)-2))*2;
if (alpha<=sig)
sprintf(['Pearson correlation between r and vx-vy is statistically significant (p=',num2str(alpha),')\n'])
else sprintf(['Pearson correlation between r and vx-vy is not statistically significant at p=',num2str(sig),'\n'])
end

function [r,nij,nx,ny,str]=nominalcorr(sel,x,y)
m=max(size(x));
nx=unique(x); p=max(size(nx));
ny=unique(y); q=max(size(ny));
for k=1:p
for j=1:q
nij(k,j)=0;
for i=1:m
if ((x(i)==nx(k)) & (y(i)==ny(j))) nij(k,j)=nij(k,j)+1; end
end; end; end
for k=1:p
ni(k)=sum(nij(k,:));
end
n=sum(ni);
for k=1:q
nj(k)=sum(nij(:,k));
end
wsquare=0;
for k=1:p
for j=1:q
wsquare=wsquare+nij(k,j)^2/(ni(k)*nj(j));
end; end
wsquare=n*(wsquare-1);
switch (sel)
  case 1
        r=var(nij(:))/(var(nij(:))+mean(nij(:))); str='association coefficient';
  case 2
        r=sqrt(wsquare/(wsquare+n)); str='contingency coefficient';
  case 3
        r=sqrt(wsquare/(n*sqrt((p-1)*(q-1)))); str='contingency coefficient I';
  case 4
        r=sqrt(wsquare/(n*max(p-1,q-1))); str='contingency coefficient II';
  case 5
        r=sqrt(wsquare/(n*min(p-1,q-1))); str='contingency coefficient III';
end
```

3.2.2 Law found from statistical simulation

The results of statistical simulation indicated that for the causality of two nominal variables with *x* as independent variable and *y* as dependent variable, there is a significant Pearson correlation between nominal correlation *r* and $V_y<V_x$, and between nominal correlation *r* and V_x-V_y; $P(V_{yx}<1)<\alpha$, where α is the significant level, e.g., =0.05, 0.1, etc.

3.3 Statistical simulation for causality inference based on observed data of nominal variables

According to the law drawn above, the two rules $V_x>V_y$ and $V_{yx}>1$ can be used as the criteria for possible causality of nominal variables *x* (independent variable) and *y* (dependent variable), and the two rules $V_x<V_y$ and $V_{yx}<1$ can be used as the criteria for possible causality of nominal variables *y* (independent variable) and *x* (dependent variable).

First, assume that the nominal correlation between the two nominal variables is statistically significant. Following the principle of randomization tests (Manly, 1997; Solow, 1993; Zhang, 2007, 2010, 2011a; Zhang and Schoenly, 1999), I propose a statistical simulation method to find the causality based on the observed data of two nominal variables *x* and *y*.

In each simulation process, two phases are implemented: the first one with *x* and *y* as independent and dependent variables respectively, and the second one with *y* and *x* as independent and dependent variables respectively. In the first phase, for *x*, randomly generate *m* values of *p* types of unique values following their frequency in the data of variable *x*, and for each value in *x*, randomly generate a value of *q* types of unique values in *y* following its frequency in the data of variable *y*. Finally, calculate and record V'_x, V'_y, and V'_{yx}. Perform the same procedure in the second phase with *y* and *x* as independent and dependent variables respectively, and calculate and record V''_x, V''_y, V''_{yx}. Matlab algorithm for the the two-phase procedure is as follows:

```
for i=1:m
tem=rand();
if (tem<pni(1)) xx1(i)=nx(1); continue; end
for j=2:p
if ((tem>=pni(j-1)) & (tem<pni(j))) xx1(i)=nx(j); break;
end; end
tem=rand();
if (tem<pnj(1)) xx2(i)=ny(1); continue; end
for j=2:q
if ((tem>=pnj(j-1)) & (tem<pnj(j))) xx2(i)=ny(j); break;
end; end
end
for i=1:m
for j=1:p
if (xx1(i)==nx(j)) ids=j; break; end
end
tem=rand();
if (tem<pnij(ids,1)) yy1(i)=ny(1);
else
for j=2:q
if ((tem>=pnij(ids,j-1)) & (tem<pnij(ids,j))) yy1(i)=ny(j); break;
end; end; end
```

```
for j=1:q
if (xx2(i)==ny(j)) ids=j; break; end
end
tem=rand();
if (tem<pnji(ids,1)) yy2(i)=nx(1);
else
for j=2:p
if ((tem>=pnji(ids,j-1)) & (tem<pnji(ids,j))) yy2(i)=nx(j); break;
end; end; end
end
[rs1,nijs1,nxs1,nys1,str]=nominalcorr(sel,xx1,yy1);
[rs2,nijs2,nxs2,nys2,str]=nominalcorr(sel,xx2,yy2);
if (isnan(rs1) | (rs1==Inf) | isnan(rs2) | (rs2==Inf)) continue; end
ss=ss+1;
vyy1=mean(sum(nijs1)); vyy2=mean(sum(nijs2));
vxx1=mean(sum(nijs1')); vxx2=mean(sum(nijs2'));
vyxyx1=var(sum(nijs1))/var(sum(nijs1')); vyxyx2=var(sum(nijs2))/var(sum(nijs2'));
res(ss,:)=[vyy1 vxx1 vyxyx1 vyy2 vxx2 vyxyx2];
```

Repeat the simulation many times. Finally, calculate statististic significance values:

$$p_{xy1}=P(V'_{yx}<1)$$
$$p_{xy2}=P(V'_{y}>V'_{x})$$
$$p_{yx1}=P(V''_{yx}<1)$$
$$p_{yx2}=P(V''_{y}>V''_{x})$$

Given the statistic significance level for causality inference, σ. If $V_{yx}>1$, $V_y<V_x$, $p_{xy1}<\sigma$, $p_{xy2}<\sigma$, $1-p_{yx1}<\sigma$, and $1-p_{yx2}<\sigma$, the nominal variable for column 1 is the independent variable, and the nominal variable for column 2 is the dependent variable. If $V_{yx}<1$, $V_y>V_x$, $1-p_{xy1}<\sigma$, $1-p_{xy2}<\sigma$, $p_{yx1}<\sigma$, and $p_{yx2}<\sigma$, the nominal variable for column 2 is the independent variable, and the nominal variable for column 1 is the dependent variable.

The full Matlab codes, causalInferNomi, of the statistical simulation mthod for causality inference based on observed data of nominal variables are as follows (see supplementary material also):

```
clear
xyd=input('Input the Excel file name of raw data (e.g., xyd.xls: xyd=(dij)m×2, i=1,2,...,m; j=1,2. In the file, column 1 is for nominal variable 1 and column 2 is for nominal variable 2): ','s');
sel=input('Choose a correlation measure (1: Association coefficient; 2: Contingency coefficient; 3: Contingency coefficient I; 4: Contingency coefficient II; 5: Contingency coefficient III): ');
alpha=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., 0.01, 0.05, 0.1, etc.): ');
sim=input('Input the number of simulations (e.g., 10000): ');
xyd=xlsread(xyd);
m=size(xyd,1);
x=xyd(:,1);
y=xyd(:,2);
```

```
[r,nij,nx,ny,str]=nominalcorr(sel,x,y);
if (isnan(r) | (r==Inf))
sprintf(['The ',str,' measure is not valid. Try to use another correlation measure.\n'])
pause();
end
xx=zeros(m,1);
yy=zeros(m,1);
ss=0; ww=0;
for s=1:sim
idx=randperm(m);
idy=randperm(m);
for i=1:m
xx(i)=x(idx(i));
yy(i)=y(idy(i));
end
[rs,nijs,nxs,nys,str]=nominalcorr(sel,xx,yy);
if (isnan(rs) | (rs==Inf)) continue; end
ss=ss+1;
if (rs>r) ww=ww+1; end
end
id=0;
if ((ww/ss)<alpha)
sprintf(['There is a significant ',str,' (r=',num2str(r),') between two nominal variables (p=',num2str(ww/ss),')\n'])
id=1;
else sprintf(['There is not significant ',str,' (r=',num2str(r),') between two nominal variables (p=',num2str(ww/ss),')\n'])
sprintf(['So, causality may not exist between two nominal variables based on ',str,'\n'])
end
vy=mean(sum(nij));
vx=mean(sum(nij'));
vyx=var(sum(nij))/var(sum(nij'));
p=max(size(nx));
q=max(size(ny));
ni=sum(nij');
nj=sum(nij);
sumni=sum(ni);
pni(1)=ni(1)/sumni;
for i=2:p
pni(i)=pni(i-1)+ni(i)/sumni;
end
sumnj=sum(nj);
pnj(1)=nj(1)/sumnj;
for i=2:q
pnj(i)=pnj(i-1)+nj(i)/sumnj;
end
for i=1:p
```

```
pnij(i,1)=nij(i,1)/ni(i);
for j=2:q
pnij(i,j)=pnij(i,j-1)+nij(i,j)/ni(i);
end; end
for i=1:q
pnji(i,1)=nij(1,i)/nj(i);
for j=2:p
pnji(i,j)=pnji(i,j-1)+nij(j,i)/nj(i);
end; end
ss=0;
for s=1:sim
for i=1:m
tem=rand();
if (tem<pni(1)) xx1(i)=nx(1); continue; end
for j=2:p
if ((tem>=pni(j-1)) & (tem<pni(j))) xx1(i)=nx(j); break;
end; end
tem=rand();
if (tem<pnj(1)) xx2(i)=ny(1); continue; end
for j=2:q
if ((tem>=pnj(j-1)) & (tem<pnj(j))) xx2(i)=ny(j); break;
end; end
end
for i=1:m
for j=1:p
if (xx1(i)==nx(j)) ids=j; break; end
end
tem=rand();
if (tem<pnij(ids,1)) yy1(i)=ny(1);
else
for j=2:q
if ((tem>=pnij(ids,j-1)) & (tem<pnij(ids,j))) yy1(i)=ny(j); break;
end; end; end
for j=1:q
if (xx2(i)==ny(j)) ids=j; break; end
end
tem=rand();
if (tem<pnji(ids,1)) yy2(i)=nx(1);
else
for j=2:p
if ((tem>=pnji(ids,j-1)) & (tem<pnji(ids,j))) yy2(i)=nx(j); break;
end; end; end
end
[rs1,nijs1,nxs1,nys1,str]=nominalcorr(sel,xx1,yy1);
[rs2,nijs2,nxs2,nys2,str]=nominalcorr(sel,xx2,yy2);
```

```
if (isnan(rs1) | (rs1==Inf) | isnan(rs2) | (rs2==Inf)) continue; end
ss=ss+1;
vyy1=mean(sum(nijs1)); vyy2=mean(sum(nijs2));
vxx1=mean(sum(nijs1')); vxx2=mean(sum(nijs2'));
vyxyx1=var(sum(nijs1))/var(sum(nijs1')); vyxyx2=var(sum(nijs2))/var(sum(nijs2'));
res(ss,:)=[vyy1 vxx1 vyxyx1 vyy2 vxx2 vyxyx2];
end
pxy1=sum(res(:,3)<1)/sim;
pxy2=sum(res(:,1)>res(:,2))/sim;
pyx1=sum(res(:,6)<1)/sim;
pyx2=sum(res(:,4)>res(:,5))/sim;
if ((id==1) & (vyx>1) & (vy<vx) & (pxy1<sig) & (pxy2<sig) & (1-pyx1<sig) & (1-pyx2<sig) )
sprintf(['The nominal variable for column 1 is the independent variable, and the nominal variable for column 2 is the dependent variable (pxy1=',num2str(pxy1),', pxy2=',num2str(pxy2),', 1-pyx1=',num2str(1-pyx1),', 1-pyx2=',num2str(1-pyx2),')\n'])
elseif ((id==1) & (vyx<1) & (vy>vx) & (1-pxy1<sig) & (1-pxy2<sig) & (pyx1<sig) & (pyx2<sig))
sprintf(['The nominal variable for column 2 is the independent variable, and the nominal variable for column 1 is the dependent variable (1-pxy1=',num2str(1-pxy1),', 1-pxy2=',num2str(1-pxy2),', pyx1=',num2str(pyx1),', pyx2=',num2str(pyx2),')\n'])
else sprintf(['However, causality may not exist between two nominal variables.\n'])
end

function [r,nij,nx,ny,str]=nominalcorr(sel,x,y)
m=max(size(x));
nx=unique(x); p=max(size(nx));
ny=unique(y); q=max(size(ny));
for k=1:p
for j=1:q
nij(k,j)=0;
for i=1:m
if ((x(i)==nx(k)) & (y(i)==ny(j))) nij(k,j)=nij(k,j)+1; end
end; end; end
for k=1:p
ni(k)=sum(nij(k,:));
end
n=sum(ni);
for k=1:q
nj(k)=sum(nij(:,k));
end
wsquare=0;
for k=1:p
for j=1:q
wsquare=wsquare+nij(k,j)^2/(ni(k)*nj(j));
end; end
wsquare=n*(wsquare-1);
switch (sel)
 case 1
```

```
        r=var(nij(:))/(var(nij(:))+mean(nij(:))); str='association coefficient';
  case 2
        r=sqrt(wsquare/(wsquare+n)); str='contingency coefficient';
  case 3
        r=sqrt(wsquare/(n*sqrt((p-1)*(q-1)))); str='contingency coefficient I';
  case 4
        r=sqrt(wsquare/(n*max(p-1,q-1))); str='contingency coefficient II';
  case 5
        r=sqrt(wsquare/(n*min(p-1,q-1))); str='contingency coefficient III';
end
```

A set of theoretical data were used to validate the method above and success of the method was overally confirmed.

4 Discussion

To increase the reliability of causality inference, the size of nominal data should be large enough (i.e., $m >> max(p, q)$). 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. It should be noted that the method is not applicable to the nominal variables with full deterministic correspondence (e.g., double strand DNA).

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 statistical simulation method for causality inference of Boolean variables. Network Biology, 11(4): 263-273

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

Zhang YT, Fang KT. 1982. Introduction to Multivariate Statistics. 393-401, Science Press, Beijing, China