Measuring tendency towards mutuality in a social network

Micha Mandel

Department of Statistics, The Hebrew University, Jerusalem, Israel

Abstract

When trying to understand the structure of a social network, the number of mutual dyads is often used to describe tendency toward mutuality (TTM). Only few works, however, deal with measures of TTM and only the work of Rao and Bandyopadhyay [Rao, A.R., Bandyopadhyay, S., 1987. Sankhya 49 Ser. A, pp. 141–188] compares between the different measures. This paper presents the different measures for TTM, their interpretation and their advantages and disadvantages. We also explore the distribution of the number of mutual dyads in graphs which are distributed according to a uniform distribution over certain special spaces. The measures are demonstrated and compared on two independent real data sets. © 2000 Elsevier Science B.V. All rights reserved.

1. Introduction

Directed graphs (digraphs) are used to illustrate and to investigate social networks — when the data consist of relations between actors (people) in a group. An example, which will be demonstrated later, is friendship among children in a school-class. One of the most basic digraph characteristics is the number of mutual dyads. This graph characteristic is often used in models which describe the social structure; its interpretation is the measure of tendency towards mutuality (TTM) or symmetry in the graph. Although the number of mutual dyads is commonly used and has an interesting interpretation, only few works focus on measuring TTM. The most comprehensive work is that of Rao and Bandyopadhyay (1987), which compares many of the properties of the various measures of TTM. The present

E-mail address: smnic@mscc.huji.ac.il (M. Mandel).

1 A distinction should be made between ‘tendency towards mutuality’ and ‘symmetry’. Whereas an empty graph is symmetric by definition, we cannot measure its tendency towards mutuality. Thus, a tendency towards mutuality should be understood as symmetry in graph after controlling on several characteristics of the graph. Alternatively, tendency towards mutuality can be looked at as a characteristic of dyads rather than a graph characteristic. This characteristic is the strength of the tendency of actors in the graph to choose actors who chose them.

0378-8733/00 $ – see front matter © 2000 Elsevier Science B.V. All rights reserved.

PII: S0378-8733(00)00027-7
paper can be seen as an extension of the work of Rao and Bandyopadyay and presents some new properties of the measures dealt with by them. In particular, we show that the Katz and Powell index (Katz and Powell, 1955) should be interpreted as a standardized variable in the case where the actors are free to make any number of choices (hereafter referred to as ‘free choices case’). We also explore the distribution of the number of mutual dyads in graphs which are distributed according to a uniform distribution over certain special spaces. We show that this distribution is approximately a Poisson distribution in graphs with given outdegrees and in graphs with a given number of arcs.

After introducing some notation in Section 2, we present the indices of TTM in Sections 3–5. In Section 6, we derive the distribution of the number of mutual dyads, and in Section 7, we demonstrate and compare the indices on real data.

2. Notation

In this paper, we consider a digraph (social network) consisting of \( g \) actors. We restrict our attention to digraphs with a single 0–1 relation and we do not allow a tie from an actor to himself. The following notation is used:

\[ X = (X_{ij}) \text{ and } x = (x_{ij}) \text{ are a potential } g \times g \text{ digraph and its realization, respectively; i.e. } X_{ij} = 1 \text{ if actor } i \text{ chose actor } j \text{ (i.e. if the arc from } i \text{ to } j \text{ exists) and } X_{ij} = 0, \text{ otherwise.} \]

\[ T = T(X) \text{ represents a vector of graph statistics/characteristics and } t = T(x) \text{ is its realization. Similarly, } M = M(X) \text{ and } m = M(x) \text{ is the number of mutual dyads and its realization.} \]

\[ X_{i+} \text{ is the outdegree of actor } i \text{ (number of choices made by actor } i \text{) and } d_i \text{ its realization.} \]

Finally, \( \Omega_T = \{ x | T(x) = t \} \) is the subspace of all graphs for which \( T(x) = t \). In this work, four different \( T \)'s (and, respectively, four different subspaces of digraphs) are considered: (A) \( T \) is empty, i.e. the potential graphs are all \( (2^{g(g-1)})g \times g \) digraphs (in this case, the notation \( \Omega \) is used); (B) \( T \) is the total number of arcs in the graph; (C) \( T \) is the vector of outdegrees; (D) \( T \) is the vectors of outdegrees and indegrees. For any choice of \( T \), the index of TTM is referred to as the measure of TTM controlling on \( T \).

3. Distance indices

The simplest indices of TTM do not assume any model for the digraph distribution or structure. Instead the ‘distance’ between the number of mutual dyads in the digraph in hand and the number of mutual dyads that could occur is measured; deterministic indices measure the distance between the number of mutual dyads and the minimum value that it can attain. Probabilistic indices of TTM measure the deviation of the number of mutual dyads from its expected value under some randomness assumption.

Formally, for any choice of \( T \), the deterministic index is

\[ D_{T}^T = \frac{m - M_{\min}}{M_{\max} - M_{\min}} \quad (1) \]

where \( M_{\max} = \max\{ m | x \in \Omega_T \} \) and \( M_{\min} = \min\{ m | x \in \Omega_T \} \).
The probabilistic indices are

\[
P_{1T}^T = \frac{M - EM}{\sqrt{\text{var}(M)}}
\]

(2)

\[
P_{2T}^T = P(M \leq m)
\]

(3)

where the expectation, variance and probability are calculated under the assumption of a uniform distribution of \(X\) over \(\Omega_T^T\).

After \(T\) is chosen, the deterministic index can be interpreted as the proportion of the number of mutual dyads to their maximal potential number. The first probabilistic index is a standardized variable which measures the distance of the number of mutual dyads from its expected value in terms of standard deviations. The second probabilistic index is \(P\)-value like: one minus this index is the probability that the number of mutual dyads be greater than the actually observed number, under a uniform distribution over the potential graphs.

A comprehensive discussion of the distance indices can be found in Rao and Bandyopadhyay (1987). They consider the four subspaces of potential \(g \times g\) digraphs that are defined in Section 2. Table 1 presents the deterministic values needed to calculate the indices.

In the ‘free choice case’ C, the determination of the deterministic index is not straightforward, but has been obtained exactly, whereas in case D, only a lower bound for \(M_{\text{min}}\) and an upper bound for \(M_{\text{max}}\) are obtained (for details, see Rao and Bandyopadhyay, 1987).

Table 2 summarizes the moments needed to evaluate the probabilistic indices.

Katz and Wilson (1956) obtain the moments for the general C case (free choice case). To the best of my knowledge, there is no known method to calculate the moments of \(M\) in case D, to date. However, Snijders (1991) derives a method to sample a random graph with given marginals. The moments can then be evaluated using a Monte Carlo simulation method.

### Table 1

<table>
<thead>
<tr>
<th>Case</th>
<th>Graphs with</th>
<th>(M_{\text{min}})</th>
<th>(M_{\text{max}})</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>All graphs</td>
<td>0</td>
<td>(g(g-1)/2)</td>
</tr>
<tr>
<td>B</td>
<td>(t) arcs</td>
<td>(\max{0, t-g(g-1)/2})</td>
<td>([t/2])</td>
</tr>
<tr>
<td>C</td>
<td>Fixed outdegrees ((d_i=d\text{ for all }i))</td>
<td>(\max{0, gd-g(g-1)/2})</td>
<td>([t/2])</td>
</tr>
</tbody>
</table>

\(a\) \([/a]\) is the largest integer less than or equal to \(a\).

### Table 2

<table>
<thead>
<tr>
<th>Case</th>
<th>Uniformly distributed over graphs with</th>
<th>(E(M/T)=\mu)</th>
<th>(\text{var}(M/T))</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>All graphs</td>
<td>(g(g-1)/8)</td>
<td>(3g(g-1)/32)</td>
</tr>
<tr>
<td>B</td>
<td>(t) arcs</td>
<td>(t(t-1)/2(g(g-1)-1))</td>
<td>(\mu(1-\mu+[{(t-2)(t-3)/2}(g(g-1)-3)])</td>
</tr>
<tr>
<td>C</td>
<td>Fixed outdegrees ((d_i=d\text{ for all }i))</td>
<td>(gd^2/2(g-1))</td>
<td>(\mu(1-d(g-1))^2)</td>
</tr>
</tbody>
</table>
The calculation of the probabilistic indices denoted PI2 can be carried out using the results of Section 6.

4. The index of Katz and Powell

A simple index of TTM was suggested by Katz and Powell (1955). Katz and Powell claim that “wherever a tendency to mutuality is real and sizeable, it is unrealistic to use random models to describe and evaluate the phenomena”. Hence, they suggest a probability model for mutual choices between two actors; assuming that the number of choices made by each actor is given, he chooses other actors randomly and according to a uniform distribution. The model assumptions are

\[ P(X_{ij} = 1 | X_{ji} = 1) = P(X_{ij} = 1) + \tau P(X_{ij} = 0) \]

\[ P(X_{ij} = 1 | X_{ji} = d_i) = \frac{d_i}{g-1} \]

where \( d_i \) is the number of choices made by actor \( i \), and \( \tau \) the Katz and Powell’s index of TTM. While the first assumption (4) defines the model, the second (5) is needed for the estimation of \( \tau \).

This index is always between \(-1\) and \(1\), and can easily be shown to be the correlation coefficient between \( X_{ij} \) and \( X_{ji} \). (From the model, it follows that the correlation coefficient between choices in dyads is constant in the graph; see Rao and Bandyopadhyay, 1987).

For the case of fixed choices (\( d_i = d \) for all \( i \)), (Katz and Powell, 1955) suggest an unbiased consistent estimator for \( \tau \) using the method of moments. Their estimator is: \( \tilde{\tau}_1 = \frac{2m(g - 1) - gd^2}{gd(g - 1 - d)} \). For the ‘free choice case’, the moments estimator generalized to \( \tilde{\tau}_2 = \frac{2(g - 1)^2m - S_1^2 + S_2}{(g - 1)^2S_1 - S_1^2 + S_2} \), where \( S_k = \sum_i d_i^k \), \( k = 1, 2 \). Whenever the \( d_i \)'s do not differ much, Katz and Powell suggest using \( \tilde{\tau}_2 \) instead of \( \tilde{\tau}_1 \) replacing \( d \) by the average of the \( d_i \)'s.

Although the index is simple to evaluate and to interpret, this model has two severe disadvantages.

First, the model assumes a very unrealistic behavior for the actors on the graph. Only in a few graphs (if any) can one assume that the graph distribution has these properties.

Second, the extension of the model from a fixed number of choices to the ‘free choice case’ is problematic; Eq. (4) can be rewritten, using the definition of conditional probability as

\[ P(X_{ij} = 1, X_{ji} = 1) = P(X_{ij} = 1)P(X_{ji} = 1) + \tau P(X_{ij} = 0)P(X_{ji} = 1) \]

The LHS of Eq. (6) is symmetric in \( i, j \), thus, the RHS must also be symmetric. It is easily seen that Eq. (6) holds if and only if for all \( (i,j) \) one has \( P(X_{ij} = 1) = P(X_{ji} = 1) \). This together with Eq. (5), yields that the \( d_i \)'s must be constant in the graph; therefore Eqs. (4) and (5) cannot both hold in the ‘free choice case’. If we do not require the actor choices to be uniformly distributed given the outdegrees (i.e. we omit Eq. (5)), then, we cannot calculate the expected value of \( M \) in this general model, and cannot estimate the desirable index.
Although we just showed that using the Katz and Powell index can be justified only in rare situations, the $\hat{e}_1$ statistic can be used as an index of mutuality with a different interpretation. Using Eq. (2) and Table 2, we find that the probabilistic index in the ‘fixed choice case’ is 

$$2m(g - 1) - gd^2]/d(g - 1 - d)^{\frac{1}{2}} = \frac{\sqrt{g(g - 1)}}{2} \hat{e}_1.$$ 

Thus, the Katz and Powell index estimator is just a standardization of the probabilistic index, forcing it to lie in the interval $[-1, 1]$.

5. Indices of TTM in the $P_1$ and $P^*$ exponential models

These models measure TTM in a somewhat different manner; whereas the indices mentioned in the previous sections measures solely TTM, the purpose of the $P_1$ and $P^*$ models is to learn the whole structure of the graph and to infer the influence of TTM embedded in this structure. Loosely speaking, the previous indices measure the extent of TTM conditional on several graph statistics whereas these indices measure the influence of TTM on the structure of the graph. This point becomes clearer when we explain the indices interpretation below.

In these models, the graph probability function is assumed to have the form

$$P(x = x) = c(\theta) \exp (\theta' T(x))$$

(7)

where $\theta$ is a vector of parameters, $T$ a vector of graph statistics, $\theta'$ the transpose of $\theta$ (row vector) and $c(\theta)$ a normalizing constant.

In the $P_1$ model, the $(g-1)/2$ dyads in the graph are assumed to be independent multinomially distributed. Additionally, the odds ratio of the choices in each dyad is assumed to be constant over the graph, that is

$$P(X_{ij} = 1|X_{ji} = 1)P(X_{ij} = 0|X_{ji} = 0) = \rho$$

(8)

Under the $P_1$ model, $\rho$ is a natural choice as an index of TTM which indicates a positive TTM if $\rho > 1$ (see Holland and Leinhardt, 1981).

Like the $P_1$ model, the $P^*$ model has the form (7). However, it does not make the additional assumption that the dyads are independently distributed (see Wasserman and Pattison, 1996). After the determination of the vector $T$ (see Eq. (7)), one can estimate the vector of parameters $\theta$ using the method of Maximum Pseudo-likelihood (Strauss and Ikeda, 1990).

In order to understand the interpretation of the parameters of $P^*$ further, let $c_{ij} = \{x_{uv} | (u, v) \neq (i, j)\}$ be a realization of the graph except the arc $x_{ij}$, and write $x_{ij}^+$, $x_{ij}^-$ for the realizations of the graph generated by $c_{ij}$ and $x_{ij}$ set to be 1 and 0, respectively. One can show that (see Strauss and Ikeda, 1990)

$$P(X_{ij} = 1|c_{ij})/P(X_{ij} = 0|c_{ij}) = \exp (\theta' [T(x_{ij}^+) - T(x_{ij}^-)])$$

(9)

where component $k$ of $\Delta(x_{ij})$ is the change in component $k$ of $T$ when the arc $x_{ij}$ is changed from 0 to 1. Thus, change in component $k$ of $\Delta(x_{ij})$ by one unit, given that the other components remain unchanged, multiplies the conditional odds ratio by $\exp (\theta_k)$. (Notice that the components of $\Delta(x_{ij})$ are strongly dependent, thus, change in one component
usually causes changes in the others.) Many models include the number of mutual dyads in the vector $T$ and use $\theta_m$ (the corresponding parameter) as measurement or index of TTM.

The interpretation of $\theta_m$ can be its contribution to the likelihood using Eq. (7) (Wasserman and Pattison, 1996). Positive (negative) $\theta_m$ values assign higher probability to graphs with a large (small) number of mutual dyads (given that the rest of the vector $T$ remains unchanged). (Anderson et al., 1999) use Eq. (9) to interpret $\theta_m$ as the conditional log odds ratio; thus, if $x_{ji}=1$, given the rest of the graph, the probability that $x_{ij}=1$ is $\exp(\theta_m)$ times larger than the probability that $x_{ij}=0$.

The indices of TTM in the $P_1$ and $P^*$ models have the advantage of measuring mutuality controlling different graph characteristics; however, like other parametric models (e.g. regression), there are some disadvantages concerning model selection and estimation.

First, these indices assume a special model for the graph probability function. The indices are thus model dependent; different models have different indices, hence, to use them, one must have a very strong theoretic belief of the data structure. In order to eliminate the assumption of a special model, one can assume a family of models, usually by defining the dependence structure of the graph (see Frank and Strauss, 1986). A model from this family can be chosen according to some model selection criterion, however, no consensus on a model selection criterion exists, and a subjective decision must be made. The different results for different models can be seen for example in Wasserman and Pattison (1996), Table 6. In this table, the estimated value of $\theta_m$ changes from 1.98 in one model to 1.33 in another.

Second, even if the exact data structure is known, the obtainable indices are only estimates of the ‘true’ value. Although we expect that different estimation methods will result in the same estimates, in some cases they yield different results (see Wong, 1987). Thus, the index may depend on the estimation method used (notice that this is mainly the technical problem of which estimation method is appropriate to the problem in hand). More specifically, the method of maximum pseudo-likelihood, used in $P^*$ models, is not a standard estimation method and its performance in small graphs is questionable. Moreover, the standard errors estimates achieved by maximum pseudo-likelihood are uncertain, since the components of the pseudo-likelihood are not independent (Strauss and Ikeda, 1990; Robins et al., 1999).

The above-mentioned disadvantages are even more severe when one wishes to compare TTM between several independent graphs which could come from different models with different dependence structures. Thus, if one’s interest is solely in TTM, simpler but more easily interpretable indices should be used for comparing TTM between graphs. Examples of such are the indices discussed in Section 3.

6. The probability function of $M$

The distribution function of $M$, when the graph is uniformly distributed over special graphs spaces, has not been investigated thoroughly and is usually approximated by the normal distribution. Below, we determine the distribution of $M$ under cases A–C both in small graphs and in large ones.
6.1. Case A — all graphs

When the graph is uniformly distributed over the set of all $g \times g$ digraphs, the dyads are independent Bernoulli random variables with probability $1/4$. Hence, $M$ has the Binomial distribution $\text{Bin}(g(g-1)/2, 1/4)$ and, thus, $\sqrt{\text{var}(M)}$ convergence in law (distribution) to the standard Normal distribution, $N(0, 1)$, as $g$ increases.

6.2. Case B — graphs with a total of $t$ arcs

In this case, the probability of $m$ mutual dyads is given by

$$P(M = m) = \frac{\binom{g(g-1)/2}{m} \binom{g(g-1)/2}{m} \binom{g(g-1)/2 - m}{t-2m}}{\binom{g(g-1)}{t}},$$

where $M_{\min} \leq m \leq M_{\max}$

(10)

where $M_{\min}, M_{\max}$ are as in Table 1.

This probability can be obtained by noting that the number of graphs with $m$ mutual dyads and $t$ arcs is equivalent to the number of graphs with $m$ mutual dyads, $t-2m$ asymmetric dyads and $g(g-1)/2 - t + m$ null dyads. The number of graphs given the dyad census (the number of mutual, asymmetric and null dyads) is known to be

$$|\mathcal{D}_{\text{dyad census}}| = \frac{2^a g(g-1)/2}{m!a!n!}$$

(11)

where $|A|$ is the number of elements in A (see Wasserman and Faust, 1994). If we substitute the above mentioned values of the dyad census into Eq. (11) and rearrange, we get the numerator of Eq. (10). The denominator is the number of graphs with $t$ arcs.

The asymptotic distribution is of interest only when $t/g \rightarrow c$ as $g$ increases, thus, the average outdegree ($d$) tends to a constant. In this case, the asymptotic distribution is Poisson with parameter $c^2/2$ (see Appendix A). How good is the Poisson approximation in small graphs? Table 3 presents the maximal difference between the exact probability distribution function and cumulative distribution function of $M$ and the corresponding functions of Poisson with parameter $d^2/2$.

As can be seen in Table 3, the approximation is better when the average outdegree is small, and its quite good for moderate graphs with a small number of arcs. We point out
that the Poisson approximation can be improved using the exact mean of $M$ as the Poisson parameter (see Table 2), instead of $d_i^2/2$ (for example, the maximal difference between the cumulative distribution functions are $0.0298$, $0.0447$ and $0.0781$ for $g=20$ and $c=2$, $3$ and $5$, respectively).

6.3. Case C — graphs with given (possibly different) outdegrees

This case is very important, since many researchers ask the actors to make a fixed number of choices. The first question that a researcher might want to investigate is whether or not the mechanism of choosing is random and uniformly distributed. The answer should be given using the probability function of $M$. This probability function is, however, known only for the case $d_i=1$ for all $i$. It is obtained by Katz and Wilson (1956) and tabulated by Katz et al. (1958), up to graphs with 14 actors. They also conjecture that $M$ has an approximate Poisson distribution for large graphs. This conjecture can be verified by the following bound on the maximal deviation distance between the distribution of $M$ and the Poisson distribution. This bound also gives sufficient conditions on the outdegrees such that the asymptotic distribution is a Poisson distribution.

Let $\lambda = E(M) = \sum_{i<j}P(X_{ij} = 1, X_{ji} = 1)$ and let $Z \sim \text{Poisson}(\lambda)$. Using the Chen–Stein method (see Arratia et al., 1989), it follows that

$$\sup_{A \subset \mathbb{Z}^+} |P(M \in A) - P(Z \in A)| \leq (b_1 + b_2) \frac{1 - e^{-\lambda}}{\lambda} \leq b_1 + b_2$$

(12)

where the supremum is over all the subsets $A$ of the non-negative integers, $\mathbb{Z}^+$,

$$b_1 = (g - 1)^{-4} \sum_{i<j}((g - 1)d_i \bar{d}_i + (g - 1)d_j \bar{d}_j - d_i d_j),$$

$$b_2 = (g - 1)^{-3} \sum_{i<j} d_i d_j \bar{d}_{ij}(d_i + d_j - 2),$$

(13)

and $\bar{d}_i, \bar{d}_{i,j}$ are the averages of the outdegrees excluding actor $i$ and actors $i, j$, respectively (see details in Appendix B).

Table 4 lists the value of the bounds for graphs with fixed outdegrees.

It is clear from the table that this bound is valuable in practice only in large graphs with small outdegrees. It follows from Eq. (12), however, that whenever $b_1 + b_2 \rightarrow 0$ the asymptotic distribution is Poisson. This includes the case of fixed choices and all cases of a bounded number of choices, which are the only cases of practical importance.

It is worth noting that in order to evaluate the probabilistic index $P_{12}$ the interesting subspaces of $\mathbb{Z}^+$ are subsets with the form $A = \{z \in \mathbb{Z}^+ | z \leq m\}$. Table 5 shows that the maximal deviation distance between the two distributions, limited to these subspaces, is usually much smaller than the bound in Eq. (12).

The distribution function of $M$ in Table 5 was estimated by simulation of 1000 random graphs. This method is very simple, and is recommended, especially in small graphs. As mentioned earlier, using simulation one can also estimate the distribution function of $M$ for the more complicated case D (see Snijders, 1991).
Table 4
Upper bound on the maximal deviation distance between the exact distribution of \( M \) and the Poisson distribution — fixed outdegrees

<table>
<thead>
<tr>
<th>( d )</th>
<th>15</th>
<th>50</th>
<th>100</th>
<th>500</th>
<th>1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.057</td>
<td>0.016</td>
<td>0.008</td>
<td>0.002</td>
<td>0.001</td>
</tr>
<tr>
<td>2</td>
<td>0.374</td>
<td>0.106</td>
<td>0.052</td>
<td>0.010</td>
<td>0.005</td>
</tr>
<tr>
<td>3</td>
<td>0.987</td>
<td>0.282</td>
<td>0.140</td>
<td>0.028</td>
<td>0.014</td>
</tr>
<tr>
<td>4</td>
<td>–</td>
<td>0.530</td>
<td>0.262</td>
<td>0.052</td>
<td>0.016</td>
</tr>
<tr>
<td>5</td>
<td>–</td>
<td>0.857</td>
<td>0.424</td>
<td>0.084</td>
<td>0.042</td>
</tr>
</tbody>
</table>

Table 5
Maximal deviation distance between the exact distribution of \( M \) (fixed outdegrees) and the Poisson distribution limited to subspaces of \( \mathbb{Z}^+ \) with the form \( \{ z \in \mathbb{Z}^+ : |z| \leq m \} \)

<table>
<thead>
<tr>
<th>( d )</th>
<th>15</th>
<th>50</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0123</td>
<td>0.0156</td>
<td>0.0133</td>
</tr>
<tr>
<td>3</td>
<td>0.0640</td>
<td>0.0264</td>
<td>0.0147</td>
</tr>
<tr>
<td>5</td>
<td>0.1185</td>
<td>0.0351</td>
<td>0.0343</td>
</tr>
</tbody>
</table>

*The exact distribution of \( M \) was derived by simulation of 1000 graphs.

7. An empirical example

To illustrate the different indices, a simple experiment among children in a sixth grade class was performed. The children in this class were asked to mark on a questionnaire all their (same gender) friends in the class. The class contained 15 girls and 16 boys. One boy was missing on the day of the experiment, but some boys marked him as their friend. I decided to include him in the calculation of the distance indices and in the \( P_1 \) model his outdegree parameter was taken to be 0. The data is given in Tables 6 and 7 for the girls and boys, respectively.

Fig. 1 presents the indices on a radar graph. \( D(#) \) is the deterministic index (DI) for case \#, \( P(#) \) is the second probabilistic index (PI2) for case \#. The deterministic index in case C is the same as in case B and is, therefore, not denoted separately. The \( P(D) \) indices were calculated using Snijders simulation method (Snijders, 1991).

All indices indicate a positive TTM, but their magnitude differ. We see that the probabilistic indices vary from 2.32 standard deviations in case A to 5.64 in case C for boys, and from 4.90 to 5.64 for girls. Moreover, whereas in case A, the TTM seems to be greater for girls both when measured by the deterministic index and the probabilistic one, the other

---

*The calculation was carried out using the ZO computer program developed by Snijders (see http://stat.gamma.rug.nl/snijders/). It is based on a sample of 10,000 random graphs with marginals as in Tables 6 and 7.*
Table 6
Friendship among girls

<table>
<thead>
<tr>
<th>Girls</th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
<th>e</th>
<th>f</th>
<th>g</th>
<th>h</th>
<th>i</th>
<th>j</th>
<th>k</th>
<th>l</th>
<th>m</th>
<th>n</th>
<th>o</th>
<th>Outdegrees</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>b</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>c</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>d</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>e</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>f</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>g</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>h</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td>i</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>j</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>k</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>l</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>m</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>n</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>o</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>12</td>
</tr>
<tr>
<td>Indegrees</td>
<td>9</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>9</td>
<td>3</td>
<td>11</td>
<td>5</td>
<td>10</td>
<td>9</td>
<td>6</td>
<td>3</td>
<td>8</td>
<td>7</td>
<td>27</td>
<td></td>
</tr>
</tbody>
</table>

Table 7
Friendship among boys

<table>
<thead>
<tr>
<th>Boys</th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
<th>e</th>
<th>f</th>
<th>g</th>
<th>h</th>
<th>i</th>
<th>j</th>
<th>k</th>
<th>l</th>
<th>m</th>
<th>n</th>
<th>o</th>
<th>p</th>
<th>Outdegrees</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>b</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>c</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>d</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>e</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>f</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>g</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>h</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>i</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>j</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>k</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>l</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>m</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>o</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>p</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Indegrees</td>
<td>4</td>
<td>6</td>
<td>8</td>
<td>4</td>
<td>7</td>
<td>8</td>
<td>7</td>
<td>8</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>7</td>
<td>8</td>
<td>105</td>
</tr>
</tbody>
</table>

cases including P1 model suggest that TTM is greater for the boys. The difference as well as the ratio between the indices for boys and girls, seems to get larger as the index calculation is conditional on more components of the graph. Thus, we see that measure of TTM is rather sensitive to the index one decides to use, i.e. to the structure of the graph one limits oneself.
8. Summary

After describing several indices dealing with measuring a tendency towards mutuality, we discuss their interpretations and limitations. We have not given a rule for choosing an index. To date the most popular index of TTM is one based on the P* family of models. Usually one is interested not in mutuality alone, but in the whole graph structure; hence, a full model, which also measures mutuality, is preferable. Although in these models one can control on the graph structure and measure the ‘pure’ TTM, the problem of model selection and index interpretation, and as a consequence the difficulty to compare TTM between independent graphs, make these indices unattractive. The distance indices, discussed in Section 3, have a simple interpretation but both the deterministic and probabilistic indices are too simple; they do not fully control the graph structure. Although theoretically one can calculate the distance indices for any graph structure, in practice this task is in most cases too complicated. Rao and Bandyopadhyay (1987) recommend the distance indices, which control on outdegrees, as the preferred indices for measuring TTM. We agree with their recommendation, in cases where the researcher’s interest is solely mutuality or when a comparison of mutuality between several independent graphs is desired. However, if one wishes to investigate the influence of TTM on the structure of the graph, then a model-based index is more appropriate.
Acknowledgements

This work is part of my masters thesis. I am deeply grateful to Prof. Esther Samuel-Chan for her guidelines and helpful comments.

Appendix A. Asymptotic distribution of M in graphs with a given number of arcs

Lemma. Let \( \{Z_n\}_{n=1}^{\infty} \) be a sequence of discrete random variables on the non-negative integers. Suppose that for each \( z \in \{0, 1, 2, \ldots\} \) there exists \( N_z \) such that for all \( n > N_z \) one has \( P(Z_n = z) > 0 \).

Let \( Z \) be a random variable for which

\[
\lim_{n \to \infty} P(Z_n = 0) = P(Z = 0) \quad \text{(A.1)}
\]

and

\[
\lim_{n \to \infty} \frac{P(Z_n = z)}{P(Z_n = z - 1)} = \frac{P(Z = z)}{P(Z = z - 1)}, \quad \text{for } z = 1, 2, 3, \ldots \quad \text{(A.2)}
\]

Then, the sequence \( Z_n \) convergence in law to \( Z \) as \( n \) increases.

The proof of this lemma by induction is simple and hence omitted.

In our case, assuming that when the number of arcs in the graph, \( t \), increases the average outdegree, \( t/g \), approaches a constant \( c \), we see from Eq. (10) that

\[
\frac{P(M = m)}{P(M = m - 1)} = \frac{(t - 2m + 2)(t - 2m + 1)}{4m(g(g - 1)/(2 + m - t))} \to c^2/2m
\]

Thus, the second condition of the lemma gives exactly the ratio for a Poisson distribution with parameter \( c^2/2 \).

Using Eq. (10) again, the first condition can be shown as follows:

\[
P(M = 0) = \frac{(g(g - 1)/2)!}{(g(g - 1))!} \frac{(g(g - 1) - t)!}{(2 - t)!} = 2^{t-1} \prod_{i=1}^{t} \left( \frac{1}{2} \frac{g(g - 1) - t + i}{g(g - 1) - t + i} \right)
\]

Using two standard approximations, we get

\[
\log(P(M = 0)) \cong - \sum_{i=1}^{t} \frac{t - i}{g(g - 1) - t + i} \cong - \sum_{i=1}^{t} \frac{t - i}{g(g - 1) - t + 0}
\]

\[
= - \frac{(t^2/2) - (t/2)}{g(g - 1) - t} \to_{t/g \to c} c^2/2
\]
The limiting value is the log of $P(Z=0)$ for a Poisson random variable $Z$ with parameter $c^2/2$, and hence, the proof is complete.

**Appendix B. Bound on the deviation of the distribution of $M$ with given outdegrees from the corresponding Poisson distribution**

We use Theorem 1 of Arratia et al. (1989).

**Notation.** Let $M_{\alpha} \sim B(1, p_\alpha)$, $\alpha \in I$ be random variables. For each $\alpha \in I$ define a group of indices $B_\alpha \subseteq I$ such that $M_\alpha, M_\beta$ are independent if $\alpha \not\in B_\beta$ and $\beta \not\in B_\alpha$.

Let

$$M = \sum_{\alpha \in I} M_{\alpha}, \quad \lambda = \sum_{\alpha \in I} p_\alpha, \quad p_{\alpha\beta} = E(M_{\alpha}M_{\beta}), \quad p_{aa} = 0.$$  

Define the two quantities

$$b_1 = \sum_{\alpha \in I} \sum_{\beta \in B_\alpha} p_\alpha p_\beta,$$

$$b_2 = \sum_{\alpha \in I} \sum_{\beta \in B_\alpha} p_{\alpha\beta}.$$

**Theorem.** Suppose $0 < \lambda < \infty$ and $Z \sim \text{Poisson}(\lambda)$. Then,

$$\sup_{A \subseteq \mathbb{Z}^+} |P(M \in A) - P(Z \in A)| \leq (b_1 + b_2) \frac{1 - e^{-\lambda}}{\lambda} \leq b_1 + b_2,$$

where the supremum is over all subsets $A$ of the non-negative integers $\mathbb{Z}^+$.

The adaptation for our case is as follows:

$M_{ij} = X_{ij}X_{ji} \sim B\left(1, \frac{d_i d_j}{(g-1)^2}\right), \quad I = \{(i, j) | 1 \leq i \leq j \leq g\},$

$$M = \sum_{(i, j) \in I} M_{ij}.$$  

Let

$$I_l = \{(k, l) | i = k \text{ or } i = l\} \text{ and } B_{ij} = I_i \cup I_j.$$  

Then, the above conditions hold.

We also need expression for

$$p_{ijk} = P(M_{ij}M_{ik} = 1) = d_i (d_i - 1) d_j d_k / (g - 1)^3 (g - 2).$$
Putting all together, we get

\[ b_1 = \sum_{i<j (i,j) \in B_{ij}} \frac{d_i d_j d_k d_l}{(g-1)^4} = (g-1)^{-4} \sum_{i<j} d_i d_j \left( \sum_{k \neq i} d_i d_k + \sum_{l \neq j} d_j d_l - d_i d_j \right), \]

\[ b_2 = \sum_{i<j (i,j) \in B_{ij}} p_{ijkl} = (g-1)^{-3} (g-2)^{-1} \]
\[ \times \sum_{i<j} \left( \sum_{k \neq i,j} d_i (d_i - 1) d_j d_k + \sum_{l \neq i,j} d_j (d_j - 1) d_i d_l \right), \]

which is a different presentation of Eq. (12).

References