Heider balance, asymmetric ties, and gender segregation

Małgorzata J. Krawczyk\textsuperscript{a}, Marcelo del Castillo-Mussot\textsuperscript{b}, Eric Hernández-Ramírez\textsuperscript{b}, Gerardo G. Naumis\textsuperscript{b}, Krzysztof Kułakowski\textsuperscript{a,}\textsuperscript{*}

\textsuperscript{a} AGH University of Science and Technology, Faculty of Physics and Applied Computer Science, al. Mickiewicza 30, 30-059 Kraków, Poland
\textsuperscript{b} Instituto de Física, Universidad Nacional Autónoma de México, 04510 México D.F., Mexico

HIGHLIGHTS

- Theory of Heider balance is generalized for the case of asymmetric ties.
- The gender segregation is evaluated from the data on 37 Mexican school classes.
- A new index of gender segregation is proposed and statistically validated.

ABSTRACT

To remove a cognitive dissonance in interpersonal relations, people tend to divide their acquaintances into friendly and hostile parts, both groups internally friendly and mutually hostile. This process is modeled as an evolution toward the Heider balance. A set of differential equations have been proposed and validated (Kułakowski et al., 2005) to model the Heider dynamics of this social and psychological process. Here we generalize the model by including the initial asymmetry of the interpersonal relations and the direct reciprocity effect which removes this asymmetry. Our model is applied to the data on enmity and friendship in 37 school classes and 4 groups of teachers in México. For each class, a stable balanced partition is obtained into two groups. The gender structure of the groups reveals stronger gender segregation in younger classes, i.e. of age below 12 years, a fact consistent with other general empirical results.

1. Introduction

The gender segregation is a well-known phenomenon which shapes our social life from the preschool years up to puberty [1]. It becomes visible already between children of age less than three years. Estimations of frequency of overall pro-social attitude of children of 33 months old indicate that it increases systematically when related to the same-sex children: for boys (from 19 vs. girls to 29 vs. boys) and girls (from 17 vs. boys to 31 vs. girls) [2]. When children grow, the effect is reinforced; the proportion of time spent with the same sex to cross-sex partners increases from 3:1 for the children of 4.5 years to 11:1 for the children of 6.5 years [3]. While the segregation is well documented, its origin remains disputable; for an extensive discussion we refer the reader to Refs. [1–5] and references cited therein. In particular, cognitive factors have been distinguished as one of the possible driving forces toward the segregation [1,2]; this approach assumes that children (even very young) are able to distinguish sex of their coplayers and their own. Another possible explanation is that boys and

\* Corresponding author.
E-mail address: kulakowski@fis.agh.edu.pl (K. Kułakowski).

http://dx.doi.org/10.1016/j.physa.2015.07.027
0378-4371/© 2015 Elsevier B.V. All rights reserved.
girls have different styles of play, respectively more aggressive and more passive, and they prefer to contact with those of the same style. As none of these theories has been confirmed in experimental studies [2,4,6], the issue remains open; yet the segregation persists under different forms in adulthood [5]. What is undisputed is that it is a group effect, i.e. children use sex (or gender) as a criterion of social grouping [1]. We note that after [1], we use the words sex and gender interchangeably.

Here we are interested in a general method of detection of the gender segregation by means of data analysis. We assume that the input data are given in the form of a matrix, which contains information on preferences of agents of a given group toward other members of this group. Data in this form has been collected and analyzed many times, with the paper of Zachary on the karate club [7] as a canonical example. The new element here is that we allow for asymmetric relations, i.e. the feeling of person A about person B is not necessarily the same as the feeling of B about A. In other words, the related matrix is asymmetric, and the related networks of relations are directed.

Focusing on social relations, when designing an appropriate method of data analysis, our starting point is the model of the change of perception: the removal of cognitive dissonance [8–11]. As the outcome, the method provides the group partition into two mutually hostile parts, with friendly relations within each part. This state is equivalent to a consistent partition of each triad of agents into a pair of friends and their common enemy; alternatively, all three ties in the triad are friendly.

The state is known as the Heider balance [12,13]. Per analogy to an Ising spin glass [14], in the balanced state there is no frustrated triads; the product of bonds is positive for each triad.

Our model approach [8,10,11] applies a set of differential equations and it has been successfully validated with using selected sociological data [7,15], yet all of them contained only symmetric relations. As was demonstrated in Ref. [16], one of the reasons is that the input data are given in the form of a matrix, which contains information on preferences of agents of a given group toward other members of this group. Data in this form has been collected and analyzed many times, with the paper of Zachary on the karate club [7] as a canonical example. The new element here is that we allow for asymmetric relations, i.e. the feeling of person A about person B is not necessarily the same as the feeling of B about A. In other words, the related matrix is asymmetric, and the related networks of relations are directed.

Our model approach [8,10,11] applies a set of differential equations and it has been successfully validated with using selected sociological data [7,15], yet all of them contained only symmetric relations. As was demonstrated in Ref. [16], our method is free from stable unbalanced (so-called ‘jammed’) states, which plague discrete methods based on the Monte-Carlo scheme [17,18]. The goal was not only to obtain the final partition, but also to simulate the actual dynamics of the cognitive process. It is important to note that the interpretation of the obtained partition depends on the context. In particular, the data analyzed in Ref. [15] contain just the list of attendance of a set of persons on their meetings, and the quality of relations is represented by the correlations of this attendance. In this case, the persons involved could be ignorant about any hostility or conflict in the group. In general, the method allows only to indicate two groups, with internal relations more tight (or friendly), than between the groups.

Our goal here is to generalize our former method to the case of non-symmetric relations. The motivation is twofold. The first is obvious: basically there is no reason to expect that the relations are symmetric, and actually – as is shown below – we are going to cope with data where they are not. The second aim is again to include the psychological process of direct reciprocity, one of the main mechanisms of triggering cooperation between otherwise selfish agents [19]. With this generalized model, we are going to capture the dynamics of gender separation process, where both contacts in dyads and third-person reactions allow to reproduce a partition into two coherent groups, and where differences in patterns of behavior accepted in these groups stabilize the partition. The generalized equations are applied to the weighted data on mutual liking and aversion, collected in 37 classes in Mexican schools and four groups of teachers. The age of the respondents in the classes varied from 9 to 23. Each individual had to indicate five persons which she or he liked most, and grade them with 1–5 points; the same was asked about disliking, with negative grades. These data are presented in the form of non-symmetric matrices, one matrix per class. As a rule, these data do not provide a coherent partition of a class into two sets. This is done with the model differential equations. The outcome can be understood as the partition most close to the initial data. The analysis of the gender distribution of the obtained partition allows to indicate the classes where the gender is meaningful as a criterion of the partition. This is done by the calculation of the correlation between the proportion of males and females in the obtained groups.

The next section is devoted to the model equations. Part of this chapter is relegated to the Appendix. There we show that when the equations in their previous form are applied to non-symmetric data, they lead to jammed states, similarly to the non-deterministic algorithms [17,18]. Coming back to Section 2, we demonstrate, that the jammed states are removed in the generalized model, except the case when the rate of the new process is too small.

In Section 3 we provide the details on the collected data, which describe the relations between children in 37 school classes. Separate data sets for four groups of adults are taken as a reference point. In Section 4 we describe the numerical results of the application of the model equations to the relations between children. The analysis of the results allows to evaluate the level of the gender classification in each class. The last section is devoted to the discussion.

2. The model

In the initial formulation [8], the time evolution of an element \( x(i, j) \) of the symmetric matrix of relations was given by

\[
\frac{dx(i, j)}{dt} = G(x(i, j)) \sum_{k} x(i, k)x(k, j)
\]

where the role of the term \( G(x) = 1 - (x/R)^2 \) was to limit the values of \( x \) to a prescribed range \((-R, R)\).

The idea behind Eq. (1) is as follows. Recall that the conditions of the removal of cognitive dissonance are as follows: (i) friend of my friend is my friend, (ii) friend of my enemy is my enemy, (iii) enemy of my friend is my enemy, (iv) enemy of my enemy is my friend [13]. Let us consider a particular relation \( x_{ij} \) between individuals \( i \) and \( j \). For all their common neighbors \( k \), the relations \( x_{ik} \) and \( x_{kj} \) are either of the same or different sign. In the former case, these two relations are either both
friendly or both hostile; in both cases, the triad \((i, j, k)\) is balanced if \(x_{ij}\) is positive. Then, the positive sign of the product \(x_{ik}x_{kj}\) is an incentive to improve the relation \(x_{ij}\). Conversely, the negative sign of the same product should lead to a deterioration of the relation \(x_{ij}\). In total, the time derivative of \(x_{ij}\) is a sum of the products \(x_{ik}x_{kj}\) over all neighbors of the pair \((i, j)\).

The application of Eq. (1) to the input data, described in the next section, converts each set of relations to a stable partition of the group into two subgroups. Although most of the relations given in the form of the matrix elements are initially zero, in the final partition all of them are equal to \(\pm 1\). These non-zero values reflect the important assumption of the Heider balance: in equilibrium, no neutral relations exist [8–10].

We note that the value of \(R\) is not relevant for stable states, as \(R\) can be incorporated to the time variable when dividing the whole equation by \(R\) and using variables \(y = x/R\), \(\tau = Rt\). Another option is to put \(G(x) = \Theta(x)\Theta(1 - x)\). As was shown for instance in Ref. [8], in a generic case the time evolution drives the matrix elements \(x(i, j)\) to their limit values \(\pm R\). Then, basically Eq. (1) can be written with \(G(x) = 1\), if the time evolution is numerically halted when the limit values \(x(i, j) = \pm 1\) are attained. In Ref. [16], the factor \(G(x)\) did not appear at all, which reduces Eq. (1) to a matrix Riccati equation and allows for an analytical solution. With this condition, the proof that Eq. (1) does not produce jammed states, given in Ref. [16], is valid at least for an initial stage of the evolution.

In the case when the input matrix \(x(i, j)\) is symmetric, it remains symmetric during the time evolution. Yet, for a non-symmetric input, jammed states appear again. We made an attempt to identify them for small groups, namely \(N = 3\) and \(N = 5\), where \(N\) is the number of agents in the group. (For \(N = 4\), the number of terms on the r.h.s. of Eq. (1) is even, which makes the analysis more complex, as the terms could cancel to zero.) Namely, we ask if there are stable states which are not balanced. The condition of stability is that for each pair \((i, j)\) such that \(i \neq j\)

\[
x(i, j) = \text{sign} \left( \sum_{k} x(i, k)x(k, j) \right).
\]

The condition of Heider balance is that for each triad of different agents \((i, j, k)\)

\[
x(i, j)x(i, k)x(k, j) = 1.
\]

We should add that for non-symmetric matrices, the order of indices does matter. For example, if the choice in Eq. (3) is \(x(i, j)x(i, k)x(k, j)\), the r.h.s. of Eq. (1) is the same for \(x(i, j)\) and \(x(j, i)\), which is not our intention. Accordingly, the time derivative of \(x(i, j)\) in Eq. (1) is \(x(i, k)x(k, j)\), and not \(x(i, k)x(j, k)\).

In the Appendix we show that for \(N = 5\) nodes, the jammed states do exist. We suppose that this is true also for larger values of \(N\). Yet, we also show there that for the network of \(N = 3\) nodes, there is no jammed states.

In the generalized model proposed here, the equations are

\[
\frac{dx(i, j)}{dt} = \alpha(x(j, i) - x(i, j)) + \frac{1 - \alpha}{N - 2} \sum_{k} x(i, k)x(k, j)
\]

where \(\alpha\) is a parameter which measures the rates of two processes; the variation of the relation \(x(i, j)\) due to the direct reciprocity (rate \(\alpha\)) and the variation due to an influence of a third-person (rate \(1 - \alpha\)). The first term is new; its action is just to gradually eliminate the difference between \(x(i, j)\) and \(x(j, i)\). It is clear that under its action, the symmetry is restored, if only the value of the parameter \(\alpha\) is large enough. Our experience for the Mexican data mentioned in Section 3 as well as for random matrices is that \(\alpha = 0.2\) is sufficient to make the state symmetric. Above this value, the first term in Eq. (4) gradually vanishes in time and any stable solution is balanced [8,16]. We note that although the factor \(G(x)\) is absent in Eq. (4), the values of the matrix elements \(x(i, j)\) are kept numerically in the range \((-1, 1)\). Let us add that for \(\alpha = 1\), the time evolution reduces to a set of independent pairs \((x(i, j), x(j, i))\), which just converge, each pair to identical values \((x(i, j) + x(j, i))/2, (x(i, j) + x(j, i))/2\). The coefficient \(N - 2\) in the denominator of the second term is just to make an interpretation of \(\alpha\) less dependent on the system size. The last remark is that if the first term in Eq. (4) is substituted just by \(\alpha x(j, i)\), the balance is not attained.

In Fig. 1, we show a typical dependence of the final state on the parameter \(\alpha\); the obtained pattern is reproduced for all investigated classes, as well as for the random non-symmetric input data. For \(0.2 < \alpha < 1\), both the number of unbalanced triads and the asymmetry of the matrix elements do vanish. In all but some cases, the obtained partition of the group does not depend on \(\alpha\), if only \(\alpha\) remains in the above range. Also, we do not observe any cyclic variations of \(x(i, j)\), found in Ref. [20]; we conclude that these oscillations are not generic.

The conclusion of this section is that the new term in Eq. (4), the one proportional to \(\alpha\), allows to remove the asymmetry if only \(\alpha > 0.2\), i.e. if only the rate of the related process is sufficiently large. There, the fraction of asymmetric links is zero, as shown in Fig. 1. Simultaneously, once the equilibrium state is symmetric, the jammed states are removed. This is shown also in Fig. 1: the number of imbalanced triads is zero in the equilibrium state, except for \(\alpha > 0.9\), where the balancing process (rate \(1 - \alpha\)) is too slow.

3. The data

We conducted friendship and animosity surveys in 37 classrooms groups in the Mexico City Metropolitan area with students of different ages and levels. In each group students were asked to rate and order some of their peers in their
Fig. 1. A typical dependence of the fraction $F$ of pairs of relations which remain asymmetric ($x_{ij} \neq x_{ji}$) in the stable state (blue circles) and the number of unbalanced triads (red squares) on the parameter $\alpha$. As a rule, the former measure vanishes for $\alpha > 0.2$; so does the latter, except the range $\alpha > 0.9$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 2. The histogram of relations directed toward individuals in the collected data.

classroom in two lists. In the first list each student wrote their 5 best friends in order, starting from the best friend, and in the second list they wrote the 5 worst acquaintances or enemies, starting from the worst of them. In this way we constructed 37 networks of interlinked friendship and animosity within each group. For each person we can sum the total arrows or directed links of either friendship or animosity to get useful information. A more complex description is obtained by assigning values to each arrow depending of the position in the survey lists. For example, values can be assigned to run from 5 (to the maximum positive feeling) to −5 (the maximum negative feeling). Then a person can get, for instance, a high value as a friend in two ways. First, if he or she is chosen as an ‘average friend’ by many peers, or else, he or she is considered to be a ‘very good friend’ by just a few peers. In general, there were some students who were able to mention 5 friends, but mention less than 5 ‘foes’. Some general features were found in almost all networks, such as strong differences between friendship and animosity in the way the links are distributed. In order to compare the distributions obtained in classrooms with a random model network, we performed computer simulations to distribute 5 links randomly of two types in each group. Then, as in the study of three elementary schools in Yucatán, México [21], we found that friendship and animosity statistical distributions exhibit different behavior.

In our case we found that the friendship networks are distributed more randomly, while the animosity networks tend to be more concentrated on fewer persons. That is, the animosity links are focused more on less members, which is an interesting fact. It is important to mention that here we analyze these results from the descriptive and statistical points of view, so further work is needed to understand the causes of the general difference in the asymmetrical distribution of friendship and animosity links in the surveyed groups.

A preliminary analysis reveals that despite the conditions of the poll, the data consist of more positive than negative feelings. This effect is visible in Fig. 2.

Also, there are some differences between preferences of girls and boys. For each person, results are gathered on how he or she is liked or unliked. In Fig. 3 we show a series of histograms, where these data are gathered. The mean values and standard deviations are calculated for girls about girls, girls about boys etc. The results are close to the overall mean
4. The results

For each class, including four groups of teachers, and for each gender separately, we calculated the estimated content of males (boys, men) and females (girls, women) in two parts. Obviously, the estimated content of each gender in each part is $k/2$ ($m/2$), where $k$ ($m$) is the actual number of females (males) in a given class. The related standard deviations $\sigma$ are calculated from the distribution

$$P(k_1; k) = 2^{-k} \frac{k!}{k_1!(k - k_1)!}$$  \hspace{1cm} (5)

where $k_1$ is the number of females in the class. Accordingly, the standard deviations are $\sigma = \sqrt{k/2}$; the same formula for $P(m_1; m)$ is used for males. The actual, i.e. calculated values of $k_1$ and $m_1$ are obtained from the numerical solution of Eq. (4), with the data collected in polls as the initial values of the matrix elements. Now we are interested in the correlations within the parts of the class. This correlation can be semi-quantitatively evaluated as

$$J(p) = \frac{(k_1(p) - \langle k_1(p) \rangle)(m_1(p) - \langle m_1(p) \rangle)}{\sigma_k(p)\sigma_m(p)}$$  \hspace{1cm} (6)
Fig. 4. The coefficient $J$, calculated in Eq. (7), against the mean age of children in the class. This coefficient is a measure of how the surplus of males in one 'camp' is correlated with a surplus of females in the other 'camp' in the same class.

In Fig. 4 we show the contributions $J(p)$ to this average, as dependent on the age of children in the class. The semi-quantitative (and not just quantitative) character of this evaluation is due to the fact that the standard deviations $\sigma(p)$ are very different from one class to another. Were $\sigma(p)$ the same, $J(p)$ could be interpreted as just the contribution from the class $p$ to the correlation function between the number of girls in one group and the number of boys in another group in the same class. Yet, $J(p)$ is the same for both parts of the class, and it depends on $k$ and $m$ in the same way. In Fig. 4 we show the contributions $J(p)$ to this average, as dependent on the age of children in the class. Clearly, $J(p)$ is particularly large in some younger classes.

We can also ask how the fact of being female (male) influences being the member of this or that group. To check this, we have applied the contingency test [22,23]: for each class $p$, we construct the $2 \times 2$ contingency Table 1.

Next, we calculate the values of $X^2$ as

$$X^2 = \frac{(k+m)(k_1m_2 - k_2m_1)^2}{mk(k_1 + m_1)(k_2 + m_2)}$$

and compare it with the related tabular values for one degree of freedom (Table I.7 in Ref. [23]). In this way we test the null hypothesis that the group composition is uncorrelated with gender. Out of seven classes of children below 12 years old, we have to reject the null hypothesis with the probability 0.99 for four classes, where $J > 4$; for one class, the value of $X^2 = 6.56$ is almost at the limit 6.63 for the same probability 0.99. For the two remaining classes and all 30 classes of older children, the condition $X^2 > 6.63$ is not fulfilled.

To conclude this chapter, the numerical solution of the model equations (4) allows to obtain an equilibrium state in the Heiderian balance for each class $p$. An index $J$ is proposed to evaluate the level of the gender segregation. The results given by this index are consistent with the results on the statistical significance of the segregation.

5. Discussion

Recall that the correlation $J(p)$ is a measure of how the surplus of males in one 'camp' of a class is correlated with a surplus of females in the other 'camp' in the same class. The results shown in Fig. 4 demonstrate that most significant contributions to this correlation come from the classes of the age below 12 years. In particular, the highest two values of $J(p)$ come from two classes: in one of them, the obtained partition is just 'all 9 girls' against 'all 9 boys'. In the other, one part contains 15 girls and 1 boy, the other part 2 girls and 11 boys. Clearly, the gender criterion is crucial in both these classes. On the other
hand, in the class of fifteen-year-olds, the partition was ‘2 girls and 5 boys’ against ‘2 girls and 5 boys’, and the related \( J(p) \) is equal to zero. The outlier for the class of almost-fourteen-year-olds is the result of the partition: ‘10 girls and 17 boys’ against ‘no girls and 9 boys’. Although the obtained \( J(p) \) happens to be strongly negative, it is easy to imagine that in the group of 9 males, girls are not accepted at all; then, the criterion of gender could be vivid there as well.

To summarize, we propose a method to evaluate the gender separation in a given group. The input to the evaluation is the matrix of feelings of the group members about other group members. These data can be non-symmetric and weighted. The partition into two parts is obtained by means of the set of differential equations, namely Eq. (4). The obtained stable partition only rarely and weakly depends on the parameter \( \alpha \); the choice \( \alpha = 1/2 \) is reasonable. Further, a new index \( J \) of the gender segregation is proposed. The results obtained with the poll data presented in Section 3 show that in classes of age below 12 years, the values of \( J \) are often about four or larger. This value can serve as a threshold for the semi-qualitative criterion of the gender segregation, until it is confirmed or denied by a better statistics.

Concluding, the model dynamics of removal the cognitive dissonance is generalized here to include the direct reciprocity. The same generalization allows to apply the model to non-symmetric relations. The formalism is applied to transform the data on sympathy and/or aversion, collected in 41 groups of students and teachers, into two sets in each group, with more friendly relations within the sets than between them. For each group, we evaluated the coefficient \( J \), given in Eq. (7), as a proposed measure of gender segregation. In our data, the results indicate that the segregation is visibly stronger for the youth younger than 12 years. This conclusion is consistent with the result of psychological research, that in adolescence the preferences for same-sex friends decline [24,5], and it strengthens this statement by providing a semi-quantitative criterion.

We hope that this work may stimulate further research on segregation in bidirectional networks in different social systems.

Acknowledgments

We thank Alelí Villaverde, Sharon Valverde, Jorge Alberto Olguín and Julio César Rubén Romo for their technical help in gathering and processing the surveys, Wojciech Słomczyński and Piotr Kościelnik for helpful comments. We acknowledge the partial financial support provided by DGAPA-UNAM, México, through Grant No. IN105814. The work was partially supported also by the Polish Ministry of Science and Higher Education and its grants for Scientific Research and by the PL-Grid Infrastructure.

Appendix. On the jammed states

It is easy to prove, that for \( N = 3 \) there is no jammed states; all stable states are balanced. The proof is graphical: only for the purpose of this proof we construct a network of six nodes, each node represents a link in the social network of \( N = 3 \) agents, and each link between a pair of nodes means that the time evolution of each node depends on the other node. For \( N = 3 \), this relation is reflexive. As shown in Fig. A.5, the network is a cycle. Further, in the stable state the sign of each node is a product of the signs of its two neighbors; then positive nodes have neighbors of the same sign, while negative neighbors have neighbors of different signs. Once two positive nodes are neighbors, all nodes must be positive. The only alternative is the sequence \( (+, -, -, +, -) \), which can start anywhere. Yet, as seen in Fig. A.5, a path between \( x(i, j) \) and \( x(j, i) \) contains two nodes for each pair \( i, j \). On the other hand, three steps along the above given sequence leaves us at a node of the same sign as at the start. Concluding, the matrix \( x(i, j) \) is symmetric.

For \( N = 5 \), the related network consists of 20 nodes, and its construction is less simple. We proved by a numerical inspection, that only 96 states out of \( 2^{20} \) are stable, yet only 16 out of those 96 are balanced. The remaining 80 states are jammed, i.e. stable and unbalanced; their typology reveals that there are only four types of such states, as shown in Fig. A.6. Within each type, the states differ just in the order of labels of the nodes. Concluding, for the non-symmetric matrices \( x(i, j) \) jammed states do exist. If an initial state is chosen sufficiently close to such a state, the system is supposed to rest there.
(a) One agent is liked by all other agents, yet he/she dislikes them; the relations are unreciprocated. All remaining relations are friendly.

(b) One agent is disliked by all other agents, yet he/she likes them; the relations are unreciprocated. All remaining relations are friendly.

(c) Two agents remain out of the internally friendly group of three agents. The relations of one of the two (on the left) with the group of three are hostile, but his/her relations with the other one (on the right) are unreciprocated, and so are the relations of the latter with the group of three.

(d) Two pairs of friends are mutually hostile. The balance is destroyed by the unreciprocated relations of the fifth agent with both pairs.

**Fig. A.6.** Four patterns of stable relations where the balance is broken, for \( N = 5 \). Dots in an ellipse mean a group of friends, with the same relations with the outsiders marked by separate dots. A friendly relation of X to Y is a continuous arrow from X to Y; a dotted arrow means a hostile relation. For each pattern, the matrix of relations is shown.

**References**