Gender patterns in a large online social network

Y. Volkovich[1], D. Laniado[1], K. E. Kappler[2], and A. Kaltenbrunner[1]

[1] Barcelona Media, Barcelona, Spain
[2] Fernuniversität in Hagen, Hagen, Germany

Abstract. Gender differences in human social and communication behavior have long been observed in various contexts. This study investigates such differences in the case of online social networking. We find a general tendency towards gender homophily, more marked for women, however users having a large circle of friends tend to have more connections with users of the opposite gender. We also inspect the temporal sequences of adding new friends and find that females are much more likely to connect with other females as their initial friends. Through studying triangle motifs broken down by gender we detect a marked tendency of users to gender segregation, i.e. to form single gender groups; this phenomenon is more accentuated for male users.

Keywords: Gender, Homophily, Social Network, Data Mining

1 Introduction

It is a common believe that men are more frequently early adopters of new technologies. However, in the case of many social media websites and services women are in the vanguard. Thus, women outnumbered men by a considerable amount for most social networking sites [6, 16] with Pinterest having the largest gender inequality [22] and LinkedIn being the only exception [15]. With technology entering the mass market, women lean in and overtake males not only in spending time on social networking platforms, but also in owning gadgets or playing casual social games [3].

Differences in styles of social interactions for males and females have been documented for centuries [4]. A seminal work [20] on quantitative analysis of gender differences introduces a network terminology to describe social relations between children and evolution of these relations over time. Many of the successive studies rely on questionnaires, surveys or direct observations by adults. We refer to [18, 26, 30] for further reading on this subject.

The technological advances led to the emergence of new ways to investigate human behavioral patterns. Examples of such new tools can be the analysis of data obtained from wearable sensors (see again [26] and references therein) or the exploration of mobile [23] and online social traces. Among the first works focused on gender differences in online friendship preferences were Lewis et al. [14] for Facebook and Thelwall [28] for MySpace. A recent study [27] analyzed online social interactions in the setting of a massive multi-player online game. Gender

homophily, the tendency of individuals to bond with similar others, was also reported for interactions in Wikipedia, a community with strong female minority; a higher presence of women was found in discussions with a more positive tone [13]. Finally, in [10, 12, 25] authors studied how gender influences linguistic style of messages in Twitter, Facebook and Wikipedia.

Nevertheless there is still a lack of understanding of gender roles in online social communications. As most of the studies rely on analysis of US-based users [1] some of these findings can be less relevant in non-US contexts. Gender influence on access to information and communication technologies often varies according to local and cultural practices [5, 17]. In this work we use a complete dump of a large Spanish social networking service to present an extensive analysis of *online gender homophily*, i.e. gender preferences emerging online. Spain is among the most "social media addicted" countries in the European Union [7] with almost 75% of the Spaniards using Internet as an instrument for communication and interaction with others.

In this study we explore dissimilarities between men and women in the way they sign up to a social network platform and they make friends online. We further discuss how gender homophily observed in the offline world is translated into the case of online social communications.

2   Paper roadmap and main results

To detect the fundamental differences between male and female usage of the SNS (social networking service) under analysis, we first compare the process of building their *ego networks*, i.e. online personal networks. Of particular interest is to inspect the gender of the first friend of each user to estimate the influence of gender on the adoption of a new technology. So, our first research question is:

(RQ1) *How does gender homophily affect SNS-adoption? Do men show a preference to accept invitations from men and women from women? Do online ego-networks grow in a gender-biased way?*

In our invitation-only Spanish SNS, we find that female users in most cases join the new social platform by following invitation by another female, and they add women as their initial friends, while for male users we don't observe any strong preference.

Next, we study gender homophily in more detail by answering the the following questions:

(RQ2) *Do females and males have similar friendship networks, both in size and composition? Is there a preference for connection among same gender users?* We find that males and females are almost indistinguishable with respect to their network size. We observe a relation between user popularity and the gender of a user's friends: users having an around average number of friends exhibit gender homophily (more marked for females), while users with few friends tend to have more female friends, and users with a large number of friends have more opposite gender friends.

Finally, we inspect the effect of gender on the network structure with our third research question:

(RQ3) *How does gender affect the network structure and the formation of transitive relationships (triangles)?*

We find evidence for gender segregation, as we observe a much larger proportion of single gender triangles than expected. This result is particularly marked for male only triangles. So, while we find in general a higher homophily for women, men exhibit a higher tendency to form gender homogeneous groups.

3   Dataset description

In contrast to many recent studies on gender difference based on large-scale online data, our dataset is complete in the sense that it contains the entire friendship network. Another advantage is that the SNS under analysis is gender-balanced, i.e. the number of male and female subscribers is practically the same. This is different from many other online platforms. Finally, it is also worth mentioning that we focus on a non US-located community, a category that is underrepresented in the literature.

The dataset (see detailed descriptions in [11, 31]) is a fully anonymized snapshot of friendship connections from the invitation-only (at the time this dataset was collected) Spanish social networking service Tuenti (`www.tuenti.com`). Similar to many other popular social networking platforms Tuenti allows users to set up their profiles, connect with friends and share links and media items. Users can interact by writing messages on each other's walls. The dataset includes about 9.8 million registered users (25% of Spain's population), their bidirectional friendship links (with the temporal order of link formation), and the directed interactions (an interaction is an exchange of a wall message) generated by the users during a three months period. There are small differences between the numbers of male and female participants (see Appendix A for the exact numbers) similar to those reported in surveys [9].

4   Building social environment

Gender has been observed to play a crucial role in defining people's decisions about adopting and using new technologies. Thus, men are more driven by instrumental factors (i.e. perceived usefulness) while women are more motivated by process and social factors [29]. We examine differences in how males and females start their online social experience, i.e. how they organize their online social environment, by comparing the order in which they are making friends.

The dataset under analysis comes from an invitation-only online platform, therefore we assume that the first friend of a user is the one who invited her or him. Although some data limitations (we only have successful, i.e. accepted, invitations, and no information about unfriending) we believe in the importance of this analysis for better understanding of social media involvement mechanisms.

The first friend: We schematically draw the difference in gender for the first and second friends. In Figure 1 (left) we look at the gender of the users who successfully invited a male user to join the SNS. We observe that males sign up through the invitation sent by another male in 55% of the cases and only in 45% of the cases after the invitation by a female. The gender bias however is much more significant for female users (Figure 1 (right)): in 72% of the cases women accept an invitation to join the online platform from another woman, and just in 28% of the cases from a man. We observe a similar trend for the second friend of a female user in the case that the first friend was already a female. However if, on the contrary, the first friend of a female was a male, the probability of being the second friend as well a male rises to 42%. For male users the dependency of the genders of the first two friends is even stronger: the second friend has in almost 6 of 10 cases the same gender as the first friend.

Friendship order: We go beyond the first two friends and look at the average number of same gender friends added by users given their gender and degree. In Figure 2 we plot the average fraction of same gender friends for the kth friend of male and female users form $k = 1$ to $1\,000$ (the Tuenti friendship limit). In the same plot we also show the average fraction of female friends for all users. We find than most women, as they join the new social platform, connect primarily to their female friends, creating female dominated ego networks. Women prefer to add other female users until their degrees grow larger than 150. When they have over 150 friends they tend to connect more with males. In Section 5.1 we confirm that females with many friends have a smaller fraction of same gender friends. For men we do not observe pronounced preferences. The only observation is that at the very beginning of their online social experience, and also when they have between 50 and 200 friends approximately, males have a slight tendency to connect preferentially with other males.

To sum up, women do organize their online social environment different from men especially in the initial steps, which suggests that they are more likely to add other women as their initial friends and to try a new service or enter a new social environment following an invitation by another woman. As there are many
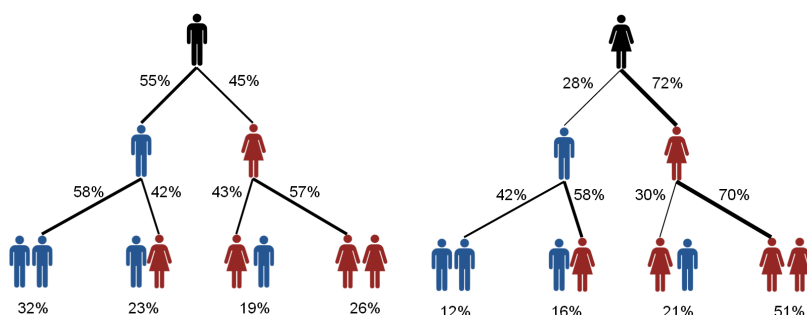

Fig. 1. Gender differences in making the first friends for males (left) and females (right).
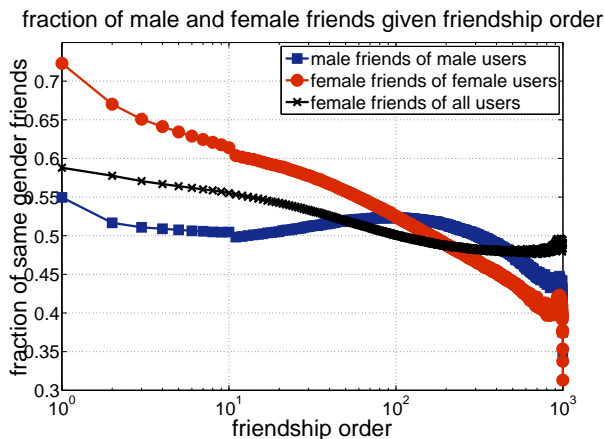
Fig. 2. Gender of the kth friend: fraction of same gender friends for male (blue squares) and female (red circles) users, and fraction of female friends of all users (black crosses) given friendship order.

different ways for users to find new friends (e.g. by using search or recommendation tools provided by SNS, through direct invitations, or by exploring friends of users' friends) further investigation is needed to explore this result.

5   Gender homophily

Exploring online friendship homophily we first find that users have just a small preference to make friends of the same gender (see detailed statistics in Appendix A). This preference is larger for females: on average male users have 82 male and 78 female friends, while females have 85 females and 76 males. The corresponding percentages are smaller in comparison to the offline world, where men are reported to have 65% and women 70% of same gender friends [24].

5.1   Gender homophily by degree

Previous work on Facebook [14] reported that males and females are almost indistinguishable with respect to their network size. In our case we also do not find any differences for degree distributions for male and female users (data not shown). However, by looking at gender ratios of users having a given degree we find that users with low (< 100) or high (> 300) numbers of friends are slightly more often females (Figure3(a)).

In Figure 3(b) we plot the ratio of female friends given the degree of a user. That is, for all users with exactly k friends, the figure shows what fraction of their friends are females, on average. We find that users with few friends tend to have more female friends; their proportion decreases with increasing degree, and
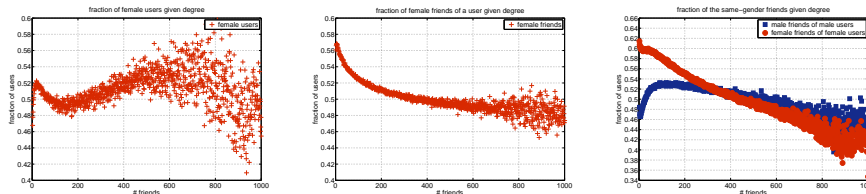
(a) Proportion of female users with a given degree.

(b) Proportion of female friends of a user given her/his degree.

(c) Proportion of same gender friends of a user given his/her degree.

Fig. 3. Gender differences given the number of friends (degree) of a user.

falls below 50% for users with more than 350 friends: users having many friends have more male friends.

To understand more deeply gender preference in friendship relationships, we also consider the fraction of same gender friends, given the degree, for male and female users separately (see Figure 3c). The figure shows, for women with few friends, a marked preference for connection with other females: around 60% for women having less than 50 friends. This preference tends to decrease with increasing degree, until women with more than 450 friends, who tend to have more male friends. For male users we observe a more balanced pattern, while we still find that users with many friends prefer to friend opposite gender users. Interestingly, males with a low number of friends also have a higher proportion of female friends. This finding is in contrast with the slight tendency of men to add other men as their initial friends, observed in Figure 2, suggesting that a preference for female friends applies only to male users having a small circle of friends (less than 25) in the SNS.

5.2 Triangle motifs

To investigate the interplay between gender and the structure of the network we next inspect gender composition of friendship triangles, i.e. triples of nodes in which each node is connected to the other two. A high presence of triangles (or a high *clustering coefficient*) is one of the key elements that distinguish social networks from other kinds of networks, such as biological or technological networks [21]. In other words, the presence of transitive relationships can be seen as a sign of a community structure, which is typical of social networks. Therefore it is particularly relevant to assess how gender affects the formation of this distinguishing pattern.

For this analysis, beyond the friendship network we consider the *interaction network*: the friendship network filtered by reciprocal interactions (i.e. keeping only connections between users who have exchanged messages on each other walls). More details about the methodology used for this analysis can be found in Appendix B.

Table 1. Proportion of triangle motifs with different gender composition (blue=male, red=female) in the friendship and interaction networks. The differences between observed (obser.) and expected proportions (shuff., calculated via reshuffling the gender of users having the same degree) are highly significant (stdv. of reshuffling $< 0.03\%$).

Type of triangle	friendship		interaction	
	obser.	shuff.	obser.	shuff.
males only	16.0%	11.6%	9.9%	6.2%
1 female, 2 males	32.5%	36.6%	24.4%	28.4%
2 females, 1 male	34.5%	38.4%	37.3%	43.3%
females only	17.0%	13.4%	28.4%	22.1%
total	3.64×10^{10}		1.24×10^{8}	

Explicit friendship triangles: In total we find more than 3.64×10^{10} triangles in the friendship network. The second and third column of Table 1 list the proportion of triangles of different composition together with the expected values based on the networks with randomly reshuffled genders. We clearly observe a much larger proportion of single gender friendship triangles than expected. In particular, although the number of female only triangles is higher, if we compare the results with the ones obtained in the reshuffled networks we find a stronger deviation for male only triangles (+38%, versus +27% for female only triangles). This indicates that the trend to form gender homogeneous groups is more accentuated for males.

Interaction triangles: When analyzing only the connections which mutually exchange messages, i.e. the interaction network, we find a striking difference between males and females, as can be observed in the two rightmost columns in Table 1. The number of female only triangles is about 3 times larger than the number of male only triangles. This difference seems high, however reshuffling shows that again we would actually have to expect an even larger disproportionality between male- and female only triangles, given that females are much more active in sending (and receiving) messages. So the tendency to form gender homogeneous groups is more marked for male users also in the interaction network. In this case the proportion of male only triangles exceeds by 60% the expected value, while the proportion of female only triangles is only 28.5% higher than expected. This indicates that male users are in general less active in the SNS, but when they interact they tend to do it in gender homogeneous groups in a much more marked measure than females.

The above results show that users do not only tend to connect preferentially with others of the same gender, but they also tend to group more by gender, and to create gender-homogeneous groups of friends. As demonstrated in [19], gender segregation is a widespread characteristic of offline social behavior. Our findings show that, in this sense, online social behavior reproduces this offline phenomenon, and that this happens more markedly for male users.

6   Conclusions

Recent studies on digital inequalities treat gender in very different ways. Some only concentrate on the influence of gender on human behavior [5], others such as Zillien [32] consider gender only as one of many variables in the emergence of digital inequalities, and yet others like boyd [2] completely ignore the gender dimension. This lack of consensus in considering gender and its influence on digital experience indicates that there are still many open questions that need to be addressed. This study is one of the first intents to shed light on emerging gender patterns in the growth of users' online personal networks.

There is growing evidence that men and women use online social platforms differently [8, 16, 27]. These differences are generally neglected when all users are treated *de-gendered* and *equally*. The analysis we present here reveals fundamental differences in how male and female users organize their online friendship networks. One of our most important findings is that females show in general a higher homophily than male users, and that this phenomenon is particularly prominent in the first steps they take in the new social environment. Women join the SNS following significantly more often an invitation from a female, and they add much more frequently other females as their initial friends.

Our findings also suggest a popularity effect, with *heterophily* characterizing users having many connections. At the same time, users having smaller circles of friends exhibit a preference for female friends irrespectively of their gender. For males, in the case when their personal network is still growing, this does not correspond to the general behavior: men tend to add slightly more frequently other men as their initial friends. For females instead we find clear evidence for homophily among women having a small or average sized personal network, as well as for women in general at their early stages in the social network (until having about 150 friends). Further research could explain whether also women who get to have large personal circles of friends (and have more male friends) still tended to exhibit homophily in their first stages.

Finally, we found evidence of homophily also in the formation of groups: the proportion of single-gender triangles is much higher than expected, reproducing the offline phenomenon of gender segregation in social behavior [19]. In contrast with the results about homophily in one-to-one friendship connections and interactions, this tendency to gender segregation is stronger for male users. Further research would be needed to investigate the gender composition of richer motifs, such as cliques and dense clusters.

Our findings show how gender affects the growth of a user's personal network and the composition and structure of friendship circles. They also unveil the importance of gender when entering a new digital social environment, and can help to understand the gender gap observed in some online communities: when females are a minority, it is less likely that other females will join, as the perceived presence of other females appears to be fundamental in the first stages.

References

1. J. Ahn. Teenagers and social network sites: Do off-line inequalities predict their online social networks? *First Monday*, 17(1), 2011.
2. d. boyd. It's complicated: the social lives of networked teens, 2014.
3. S. G. Cook. Women lead in adopting new technologies. *Women in Higher Education*, 21(2):24–25, 2012.
4. C. Darwin. *The Descent of Man and Selection in Relation to Sex*. John Murray, 1871.
5. T. Drabowicz. Gender and digital usage inequality among adolescents: A comparative study of 39 countries. *Computers & Education*, 74:98–111, 2014.
6. M. Duggan and J. Brenner. The demographics of social media users – 2012. *Pew Research Center*, 2013.
7. Eurobarometer. E-communications household survey summar. Technical report, Public Opinion Analysis, European Comission, 2010.
8. A. Hoffman. The Social Media Gender Gap. *Bloomberg Businessweek*, May, 2008.
9. Instituto Nacional de Estadistica. Mujeres y hombres en españa. Technical report, 2013.
10. D. Iosub, D. Laniado, C. Castillo, M. F. Morell, and A. Kaltenbrunner. Emotions under discussion: Gender, status and communication in online collaboration. *PloS one*, 9(8):e104880, 2014.
11. A. Kaltenbrunner, S. Scellato, Y. Volkovich, D. Laniado, D. Currie, E. J. Jutemar, and C. Mascolo. Far from the eyes, close on the Web: impact of geographic distance on online social interactions. In *Proceedings of ACM SIGCOMM Workshop on Online Social Networks (WOSN '12)*. ACM, 2012.
12. F. Kivran-Swaine, S. Brody, and M. Naaman. Effects of gender and tie strength on twitter interactions. *First Monday*, 18(9), 2013.
13. D. Laniado, A. Kaltenbrunner, C. Castillo, and M. F. Morell. Emotions and dialogue in a peer-production community: the case of Wikipedia. In *Proc. WikiSym*, 2012.
14. K. Lewis, J. Kaufman, M. Gonzalez, A. Wimmer, and N. Christakis. Tastes, ties, and time: A new social network dataset using Facebook.com. *Social Networks*, 30(4):330–342, 2008.
15. R. MacManus. Study: Women outnumber men on most social networks. `http://readwrite.com/2008/07/29/social_networks_women_outnumber_men`, 2008. Report by *Rapleaf*.
16. M. Madden, A. Lenhart, S. Cortesi, U. Gasser, M. Duggan, A. Smith, and M. Beaton. Teens, social media, and privacy. Technical report, Pew Internet Research, 2013.
17. G. Magno and I. Weber. International gender differences and gaps in online social networks. In *Social Informatics - Third International Conference, SocInfo 2014*, 2014.
18. M. McPherson, L. Smith-Lovin, and J. M. Cook. Birds of a feather: Homophily in social networks. *Annual review of sociology*, pages 415–444, 2001.
19. C. M. Mehta and J. Strough. Sex segregation in friendships and normative contexts across the life span. *Developmental Review*, 29(3):201–220, 2009.
20. J. L. Moreno. *Who shall survive? Foundations of sociometry, group psychotherapy and socio-drama*. Beacon House, 1953.
21. M. E. Newman and J. Park. Why social networks are different from other types of networks. *Physical Review E*, 68(3):036122, 2003.

22. R. Ottoni, J. P. Pesce, D. Las Casas, G. Franciscani Jr, W. Meira Jr, P. Kumaraguru, and V. Almeida. Ladies First: Analyzing Gender Roles and Behaviors in Pinterest. In *Proc. ICWSM*, 2013.
23. V. Palchykov, K. Kaski, J. Kertész, A.-L. Barabási, and R. I. Dunbar. Sex differences in intimate relationships. *Scientific Reports*, 2, 2012.
24. H. M. Reeder. The effect of gender role orientation on same-and cross-sex friendship formation. *Sex Roles*, 49(3-4):143–152, 2003.
25. H. A. Schwartz, J. C. Eichstaedt, M. L. Kern, L. Dziurzynski, S. M. Ramones, M. Agrawal, A. Shah, M. Kosinski, D. Stillwell, M. E. Seligman, et al. Personality, Gender, and Age in the Language of Social Media: The Open-Vocabulary Approach. *PloS one*, 8(9):e73791, 2013.
26. J. Stehlé, F. Charbonnier, T. Picard, C. Cattuto, and A. Barrat. Gender homophily from spatial behavior in a primary school: a sociometric study. *Social Networks*, 35(4):604–613, 2013.
27. M. Szell and S. Thurner. How women organize social networks different from men. *Scientific reports*, 3, 2013.
28. M. Thelwall. Social networks, gender, and friending: An analysis of MySpace member profiles. *Journal of the American Society for Information Science and Technology*, 59(8):1321–1330, 2008.
29. V. Venkatesh and M. G. Morris. Why don't men ever stop to ask for directions? Gender, social influence, and their role in technology acceptance and usage behavior. *MIS quarterly*, pages 115–139, 2000.
30. J. M. Vigil. Asymmetries in the friendship preferences and social styles of men and women. *Human Nature*, 18(2):143–161, 2007.
31. Y. Volkovich, S. Scellato, D. Laniado, C. Mascolo, and A. Kaltenbrunner. The Length of Bridge Ties: Structural and Geographic Properties of Online Social Interactions. In *Proc. ICWSM*, 2012.
32. N. Zillien. *Digitale Ungleichheit*. Springer, 2008.

A   Detailed statistics for gender homophily

Table A1 reports the number of male and female users in the Tuenti SNS. Quanti-
ties are shown for both the whole dataset and the filtered dataset (i.e. considering
only users having more than 10 friends).

Table A2 shows the average number of friends for male and female users,
broken down by gender. In Figure A1 we plot the complete distribution of the
percentage of same gender friends for users with more than 10 friends. We observe
that the red bars are more shifted to the right, indicating greater homophily for
females.

Table A1. Number of users in the dataset broken down by gender. The second column
shows these numbers for users with more than 10 friends.

# users	total	> 10 friends
male	4 899 659	3 269 611
female	4 784 975	3 350 189

Table A2. Basic friendship statistics by gender (all averages are taken over users with
more than 10 friends) together with 25% and 75% quantiles.

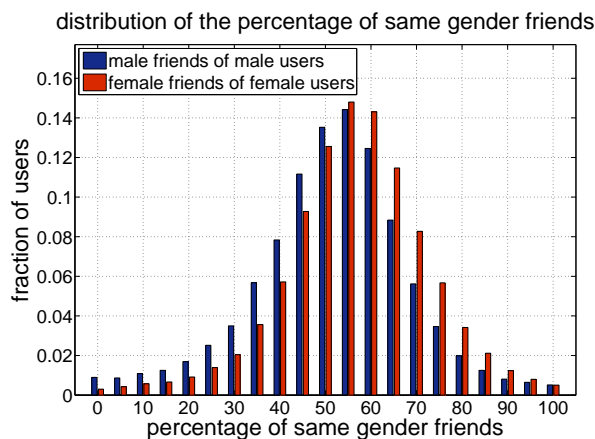| friends	avg # male	avg # female	avg % same gender
male	82[20, 116]	78[19, 106]	51.48%
female	76[15, 104]	85[23, 122]	56.46%

Fig. A1. Distribution of the percentage of the number of same gender friends for users
with more than 10 friends

B   Methodology for assessing gender homophily in transitive relationships (triangle motifs)

To explore the gender composition of friendship triangles we first focus on the entire friendship network and then restrict our analysis only to friends for which we observe reciprocal interactions. In the latter case we only consider a connection between two friends if they have sent to each other at least one wall message. We call this filtered network the interaction network. To construct it we use the information of all wall message exchanges over a period of 3 months. The resulting network is composed of 2 247 992 male and 2 521 200 female users. The number of connections for both networks, broken down by gender, is reported in Table B1. Note that for this analysis we did not filter out users having less than 10 connections. The higher number of connections involving females in the interaction network indicates that women are much more active than men in sending (and receiving) wall messages in the SNS.

Table B1. Number of connections in the friendship network and in the network of reciprocal interactions, broken down by gender.

# connections	male-male	female-female	mixed
# friendship	135 064 946	143 740 462	256 894 050
# interactions	12 236 165	22 698 114	27 346 769

There are four possibilities for the gender composition of the triangles: 3 females, 3 males, 1 male and 2 females, or 2 males and 1 female. In case of a perfectly gender balanced network, one could expect, using the binomial distribution, to have exactly 12.5% male-only triangles, 12.5% female-only triangles, and 37.5% of the triangles in each of the two mixed triangle possibilities. However, the numbers of males and females in the networks are not equal, and more importantly, the degree distributions are not equal. Females have more connections, especially in the interaction network, and this leads to a higher number of triangles involving females.

To compensate for the bias we assess how the results we observe differ from the results one should expect given the user composition of the networks. We produce randomized equivalents of our networks by re-shuffling user genders. To maintain the same gender proportions, and the same degree distribution for each gender, we randomly re-shuffle the gender of all users having the same degree. The resulting networks have the same structure and the same number of connections involving males and females as the original network. Comparing the proportion of triangles observed in the real networks with the average proportion obtained in 10 of these reshuffled networks, we are able to assess how gender influences the formation of transitive relationships. The results presented in Section 5.2 are highly significant: the standard deviation of the values observed for the reshuffled networks is smaller than 0.03%.