SCIENTIFIC REPORTS

OPEN

Dynamics of Disagreement: Large-Scale Temporal Network Analysis Reveals Negative Interactions in Online Collaboration

Milena Tsvetkova, Ruth García-Gavilanes & Taha Yasseri

Disagreement and conflict are a fact of social life. However, negative interactions are rarely explicitly declared and recorded and this makes them hard for scientists to study. In an attempt to understand the structural and temporal features of negative interactions in the community, we use complex network methods to analyze patterns in the timing and configuration of reverts of article edits to Wikipedia. We investigate how often and how fast pairs of reverts occur compared to a null model in order to control for patterns that are natural to the content production or are due to the internal rules of Wikipedia. Our results suggest that Wikipedia editors systematically revert the same person, revert back their reverter, and come to defend a reverted editor. We further relate these interactions to the status of the involved editors. Even though the individual reverts might not necessarily be negative social interactions, our analysis points to the existence of certain patterns of negative social dynamics within the community of editors. Some of these patterns have not been previously explored and carry implications for the knowledge collection practice conducted on Wikipedia. Our method can be applied to other large-scale temporal collaboration networks to identify the existence of negative social interactions and other social processes.

In the last few decades, network science has significantly advanced our understanding of the structure and dynamics of the human social fabric[1]. Much of the research, however, has focused on positive relations and interactions such as friendship and collaboration[2,3]. Considerably less is known about networks of negative social interactions such as distrust, disapproval, and disagreement. Although such interactions are not as common, they strongly affect people's psychological well-being, physical health, and work performance[4–8].

Previous work has investigated social networks with negative relations mainly from the perspective of structural balance theory[9–11]. According to this theory, the structure of social networks with both positive and negative relations, also known as "signed networks," should follow principles such as "the friend of my friend is my friend" and "the enemy of my friend is my enemy." Empirical research has found evidence that balanced triads are indeed dominant[12–14], that the global network structure tends to be balanced[15,16], and that networks evolve to increase their balance[17]. While this research illuminates how negative social relations interact with positive social relations, we still know very little about the structure and dynamics of negative interactions. For example, we have limited knowledge regarding the extent to which interaction mechanisms that we have observed for positive interactions, such as direct reciprocity and generalized, or "pay-it-forward" reciprocity, transfer to negative interactions.

Motivated by these gaps in our knowledge, we study a large online collaboration community and analyze the timing and configuration of sequences of contributions to identify patterns of negative social interactions among users. In such communities, users sometimes undo or downrate contributions made by other users. Most often these actions intend to maintain and improve the collaborative project. However, it is also possible that these actions are social in nature. Although we cannot tell whether a particular content-related action is dictated by hostility towards another user, we can study the dynamic patterns of such actions in aggregate to find if social processes occur in the community and if so, the form they take. This is what we do here.

We focus on six interactions, each of which is a sequence of two content-related actions between two or three users. We analyze the six interactions as temporal "motifs"[18,19]. We compare the observed frequency and

Oxford Internet Institute, University of Oxford, Oxford OX1 3JS, United Kingdom. Correspondence and requests for materials should be addressed to T.Y. (email: taha.yasseri@oii.ox.ac.uk)

Language	Directed, multi-link			Directed, single-link			Undirected
	Num. nodes	Num. links	Reciprocity	Out-degree assortativity	In-degree assortativity	Avg. path length	Avg. clustering
English	701169	3516109	0.183	−0.082	0.031	4.01	0.055
Spanish	58365	265291	0.202	−0.096	−0.074	3.49	0.149
German	38334	151414	0.224	−0.074	0.054	4.08	0.034
Japanese	32314	150265	0.230	−0.097	−0.073	4.16	0.042
French	31727	122374	0.191	−0.126	0.026	3.92	0.051
Portuguese	31287	159963	0.226	−0.117	−0.042	3.32	0.199
Chinese	19359	114459	0.260	−0.105	−0.099	3.29	0.145
Hebrew	11778	79989	0.264	−0.089	−0.047	3.11	0.283
Arabic	7365	28619	0.186	−0.161	−0.047	3.24	0.197
Hungarian	6114	23657	0.191	−0.154	0.026	3.38	0.126
Persian	6013	34879	0.294	−0.164	0.160	3.10	0.264
Czech	4542	19498	0.239	−0.103	−0.030	3.27	0.160
Romanian	4019	12974	0.187	−0.186	−0.209	3.04	0.179

Table 1. Descriptive statistics for the 13 Wikipedia revert networks studied.

dynamics of the motifs to those generated by a null model without any systematic clustering of actions in time for individuals but with equivalent daily patterns and community structure. In addition, we study the status effects associated with the motifs, where we use the contributor's volume of activity on the collaborative platform as a proxy for status. We confirm that the interaction is social if the corresponding motif exhibits time patterns that significantly differ from those observed in the null model and if the status effects fit with our expectations based on social psychology and results from previous studies. To present a richer picture of the observed patterns of interaction, we also look for prominent cultural effects associated with the motifs.

Our data come from Wikipedia, the largest free-access online encyclopedia. There has been much research on the coverage, quality, and biases of information provided by Wikipedia[20–22] and the controversy of topics and articles on it[23,24]. Wikipedia and its community of editors have also been extensively studied to investigate fundamental social processes. For example, using the Wikipedia community as a test case, previous research has shown that social rewards improve individual effort[25], language choice reveals power relations[26], success breeds success[27], and conflict can be both productive and counter-productive[28–30].

Wikipedia is edited by millions of volunteers. Usually, editors add text to articles, polish the existing text, or correct minor errors. Sometimes, however, they revert other editors–they undo other editors' contributions by restoring an earlier version of the article. The option to "revert" an edit is a specific feature of the Wikimedia software that was originally implemented to cope with vandalism. Reverts are technically easy to detect regardless of the context and the language and hence, they enable analysis at the scale of the whole system.

Although reverts on Wikipedia are intended to improve the encyclopedia's content, previous research has acknowledged that they could also imply negative social interactions. According to one perspective, reverts constitute maintenance work and imply conflict and coordination costs[31,32]. According to another perspective, the number of words added, deleted, and restored can be used to visualize the bipolarity of the network of editors behind an article, which in turn is associated with the perceived controversy of the article[33]. In fact, information on reverts can be used to construct a measure of article controversy and study editorial wars in relation to topics and articles[34–36]. Further, the time-series of reverts and non-revert edits can be analyzed to infer states of conflict and cooperation over the history of an article[37]. Reverts can also be combined with other actions with negative valence on Wikipedia, such as voting against an editor's promotion to admin level, to infer a signed network of relations[14]. Finally, Wikipedia reverts have been used to develop formal mathematical models of opinion dynamics explaining conflict in collaborative environments in general and beyond the Wikipedia platform[38,39].

Similarly to these previous studies, we acknowledge that a revert is not necessarily socially motivated but reverts consistently occurring in certain patterns could be signs of negative social interactions. Going beyond most existing research, we analyze the dynamics of the entire networks of reverts by considering various temporal motifs. Previous work has analyzed the distribution and structure of editing sequences and editing motifs on Wikipedia[40,41] and demonstrated the potential of this method to reveal unknown patterns in online collaboration. We here apply this method to reverts only in order to reveal patterns of negative interactions on Wikipedia. We are also the first to analyze data of reverts on multiple language editions of Wikipedia as social networks. The networks we analyze comprise about 4.7 million reverts over the period January 2001–October 2011 among the editors of 13 different language editions of Wikipedia: English, Spanish, German, Japanese, French, Portuguese, Chinese, Hebrew, Arabic, Hungarian, Persian, Czech, and Romanian. The evidence we find suggests that certain social processes interfere with how knowledge is negotiated on Wikipedia. The approach we assume can be applied to other online collaboration networks and content communities to identify the presence of similar interferences.

Results

To build the revert networks, we create time-stamped directed links going from the reverters to the reverted users. The aggregate revert networks in different language editions vary in size from 4,019 to 701,169 nodes (Table 1).
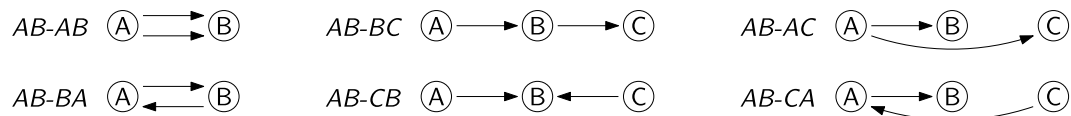
Figure 1. The six two-event temporal motifs studied. An arrow from *A* to *B* signifies that *A* reverted the content contributed by *B*. An arrow on top or to the left indicates an earlier event than an arrow on the bottom or to the right.

All networks have average clustering and reciprocity that are significantly larger than expected in a simple random graph with the same size and density (these results were confirmed in exponential-random graph models[42,43] not reported here). The giant components in the networks have short path lengths in the range 3.04–4.16. The out-degree and in-degree distributions are fat-tailed, consisting of a few editors with many reverts and many editors with only a few reverts (Supplementary Fig. S1). Further, the out-degree assortativity in the directed networks is significantly negative, implying that users who revert frequently tend to revert users who revert rarely. There is no evidence for in-degree assortativity.

Our analysis focuses on temporal motifs, which are classes of event sequences that are similar not only in the topology but also in the temporal order of the events[44]. In particular, we analyze the two-event temporal motifs in which after *A* reverts *B*, *A* reverts *B* again (*AB-AB*), *B* reverts *A* back (*AB-BA*), *B* reverts *C* (*AB-BC*), *C* reverts *B* (*AB-CB*), *A* reverts *C* (*AB-AC*), and *C* reverts *A* (*AB-CA*), as illustrated in Fig. 1. To identify the motifs, we look at every revert and identify if and when a response of these six forms occurs, restricted to a time window of 24 hours. The response may happen either in the same or in a different article. Further, the response may occur immediately after the original revert or alternatively, the reverter and the reverted may be involved in other reverts in-between the original revert and the response.

Figure 2 illustrates the identified motifs in an example time-snippet of the English-language network. If the actions in the temporal motif are socially motivated, then the motif represents a particular social interaction. For example, if *B* reciprocates a revert by *A* (*AB-BA*), this could indicate "self-defense" (undoing the "attack") or "retaliation" (responding to the "attacker" with the same); if *B* reverts someone else after being reverted (*AB-BC*) then this could be a case of "pay it forward" retaliation. Our method does not allow us to establish the exact meaning of the social interaction as it is extremely hard to make causal claims with observational data. What we instead do is to look for evidence that the interaction is social. We assume if this is the case, the temporal motif will be common, occur fast, be associated with social status effects, and possibly, be related to prominent cultural differences between editors working in different languages. Thus, our analyses test against the null hypothesis that the six motifs we observe are not socially motivated and that they do not occur more often than random chance and that they are not associated with status effects as predicted by social psychology theories and previous research.

Thus, to establish evidence for social interactions, we need appropriate null models for comparison. First, if the motifs occur intentionally and systematically, they should exhibit time patterns that differ from any random collection of pairs of reverts. Typically, in motif analysis, the null model is a suitably randomized version of the empirical data[18]. The null model we use preserves the network structure and major temporal features such as burstiness but shuffles the timestamps of events within individuals and within a window of 24 hours. The shuffle occurs with these restrictions to preserve the general patterns of individual activity and to identify each social interaction against this expected sequence of events.

Second, if the motifs represent social interactions, they should exhibit social status patterns that differ from those associated with any other pairs of reverts. We rely on common social psychology theories and previous research to derive expected status effects for most of the interactions we study and we test them below. To measure status, we use the base-ten logarithm of the number of edits the editor has completed by the time of the revert under question. This measure has a number of advantages to alternative operationalizations of status in the community of editors. The number of edits gives us a well-balanced continuous measure of experience and seniority that is easy to compare across different language versions of Wikipedia and that can be used over the entire ten-year period of observation. Importantly, the number of edits has been shown to be the strongest predictor for promotion to administrator status[45].

Here, we visualize results only for the largest network we study, the English language Wikipedia; the results for the other 12 networks are reported in the Supplementary Information. Since we test multiple hypotheses on multiple large networks, we approach the significance of results with more scrutiny. We take a conservative approach and only interpret the results that are consistent in direction and significance across different languages and operationalizations (see *Materials and Methods*). Most of the differences observed by language are not robust to different operationalizations and hence, apart from the few cases mentioned below, we do not attribute different results by language to cultural differences.

AB-AB. In our data, the *AB-AB* motif is more common and with a shorter time interval between the two actions than expected (Fig. 3a; see also Supplementary Fig. S2, Supplementary Tables S1 and S2). In general, more active and experienced editors do more reverts and hence, the reverters tend to have higher status than the reverted user. The same holds for the *AB-AB* motif, without any significant deviations for most languages (see Supplementary Fig. S8 and Supplementary Table S3). This result is somewhat surprising because we expected that individuals that are closer in status and thus, in a sense, status rivals, will be more likely to revert each other repeatedly[46].

AB-BA. The *AB-BA* motif occurs more commonly and the response rate (the time lapse between *AB* and *BA*) is shorter than what the null model generates (Fig. 3b). However, for the Chinese and Japanese languages
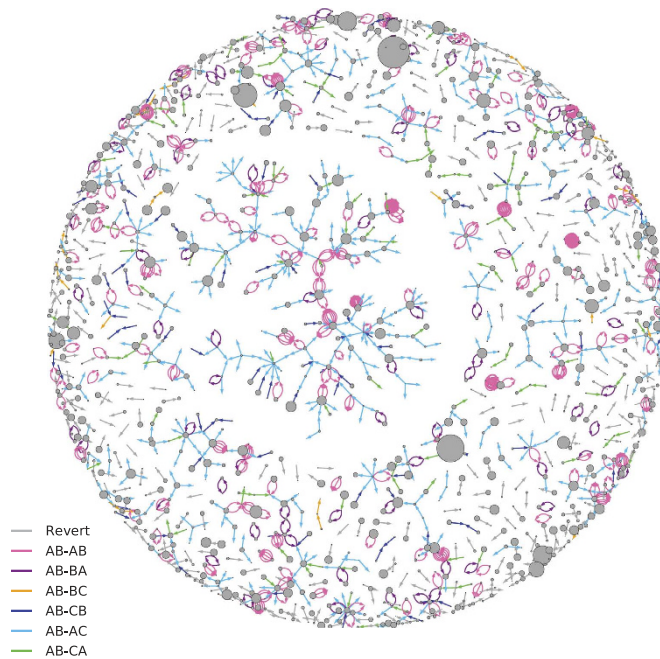
Figure 2. An example network of all reverts done in the English language Wikipedia within one day (January 15, 2010). The six temporal motifs that we investigate are color-coded. A link points from the reverter to the reverted user. The area of the nodes corresponds to the total number of edits by the user. A revert can be counted as part of more than one motif but for presentation purposes, we show only one, with the following priority: *AB-AB, AB-BA, AB-AC, AB-CA, AB-CB, AB-BC*. The network is drawn using the Kamada-Kawai force-directed algorithm.

the opposite is true–the *AB-BA* motif is not more common and the response time is significantly longer than expected (Supplementary Fig. S3). The revert networks for these two languages do not differ in any other structural aspect from the networks of the other editions. Further, no such deviations are observed in any of the other interactions that we study. Since there is no specific editorial rule that may drive these deviations, we attribute them to cultural differences.

Social comparison theory states that people strive to gain accurate self-evaluations and as a result, they tend to compare themselves to those who are similar[47]. But since focus on relative performance heightens feelings of competitiveness, rivalry is stronger among similar individuals[48]. Indeed, we find that for all languages, the users involved in the *AB-BA* motif tend to be closer in status than expected (Supplementary Fig. S9). In fact, the average difference in status is negative, meaning that *A* tends to have lower status than *B*. Additional analyses show that the status effect is driven by both equal-status and low-status reverters (Fig. 4).

We also found that the motif is more likely to occur in the same article compared to a baseline built by randomly pairing each node's reverts (Supplementary Table S4). This suggests that the interaction is largely driven by "self-defense", whereby when editor *A* reverts *B*, *B* reverts that revert. Nevertheless, up to 50% of the back-reverts occur in different articles.

***AB-BC*.** The *AB-BC* motif does not occur more commonly than expected and when it happens, the response time is longer (Fig. 3c). The literature suggests that individuals pay it forward to lower-status individuals[49,50] but this is not the case here. The only status effect we observe is that *B* is likely to revert someone else if *B* was reverted by a lower-status user (Supplementary Fig. S10 and Supplementary Table S3).

***AB-CB*.** The *AB-CB* motif is not more common (Fig. 3d) although when it happens, it appears slightly faster (except for the Arabic, Chinese, and Japanese Wikipedias; Supplementary Table S2). In the *AB-CB* motif, *A* and *C* have a leader-follower relationship[51] and from organizational theory we know that rational actors imitate each other when they want to maintain relative competitive position[52]. We thus expected that *C* is likely to be competing for status with *A*, implying that *C* should have slightly lower status than *A*. However, this does not appear to be the case (Supplementary Fig. S11 and Supplementary Table S3). In general, we find no significant status effects associated with the *AB-CB* motif.

***AB-AC*.** The *AB-AC* motif is rare in our data, and the time interval between two successive actions is longer than expected (Fig. 3e). The reverts in this motif are executed by users with high status on users with low status (Supplementary Fig. S12). Low-status users do not participate in such reverts (Fig. 4c). We expected these status effects as we know that more senior Wikipedia editors tend to scout out for vandals and watch out for edits made by newcomers. Thus, we do not find evidence for systematic serial attacks but rather for senior editors "keeping the streets clean."

***AB-CA*.** The *AB-CA* motif is more common than expected, although with a response rate as fast as expected. (Fig. 3f). On average, *C* is significantly lower in status than both *A* and *B* (Supplementary Fig. S13). One
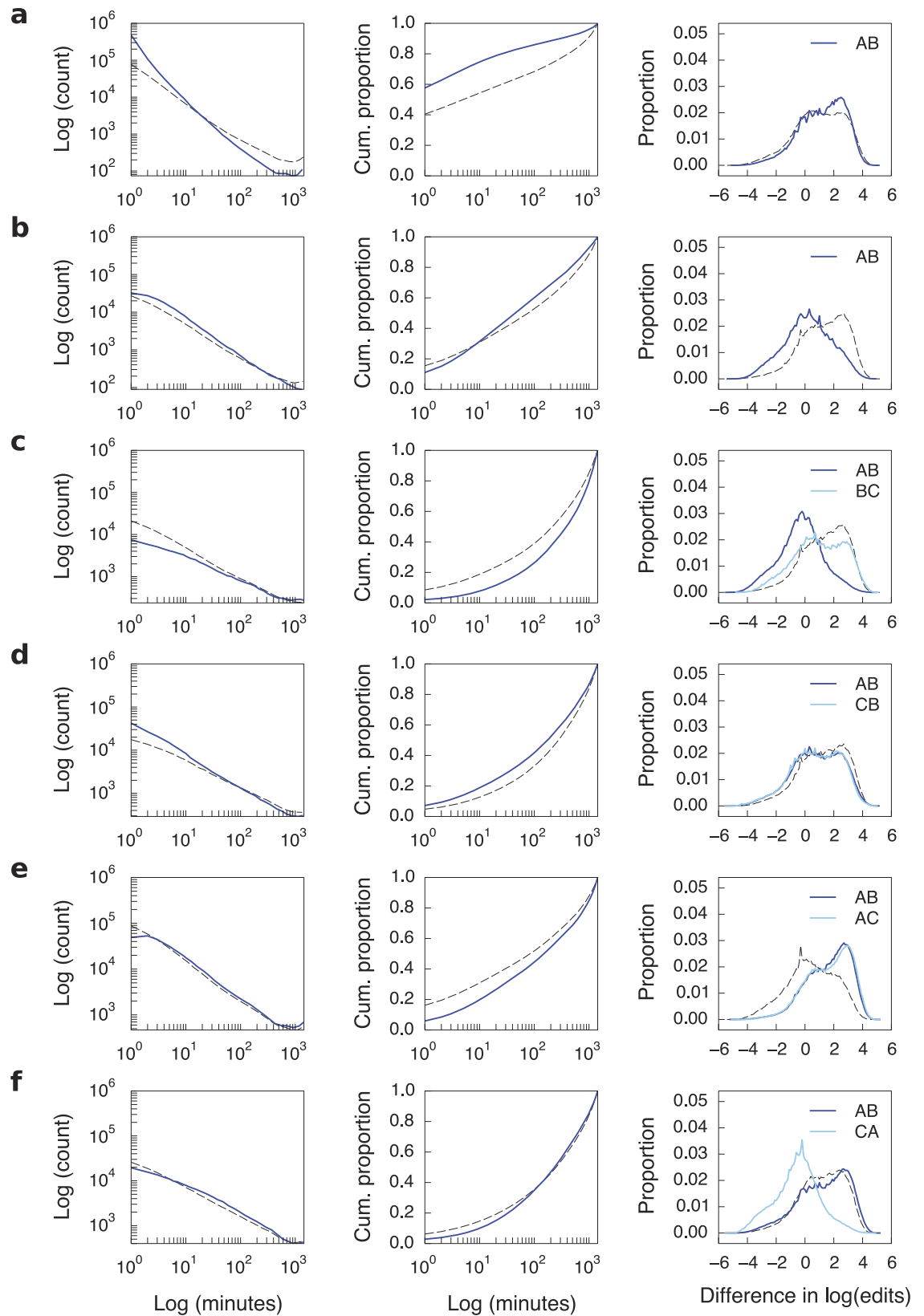
Figure 3. The frequency, rate, and status difference for the six temporal motifs. Results are shown for the English language version of Wikipedia. The dashed line shows the mean expectation from the null model (first two columns; error bars are too small to be shown) or all other reverts (third column). The results on frequency are quantified using z-scores for counts. The results on rate are quantified using multiple measures: Kolmogorov-Smirnov test statistic, mean time, and skewness. The results on status are quantified using regression analyses with standard errors adjusted for clustering within reverter and reverted. (**a**) The *AB-AB* temporal motif occurs more commonly and with a faster response rate than expected. The difference in status

5

between the reverter and the reverted user is significantly larger for English but this result is not significant for the majority of languages. (**b**) The *AB-BA* motif is more common and occurs at a faster rate than expected. The results are consistent for all languages except Japanese and Chinese. Regardless of language, the editors involved have significantly smaller status differences and in fact, on average, *B* has smaller status than *A*. (**c**) The *AB-BC* motif does not occur more commonly and has slower response rate than expected. The difference in status between *A* and *B* is significantly smaller and in fact, negative. (**d**) The *AB-CB* motif is not more common than expected but usually occurs at a faster rate. There is no evidence for status effects, as the results observed here for English Wikipedia are not consistent across the other languages. (**e**) The *AB-AC* motif does not occur more commonly and is in fact slower than expected. The difference in status between the reverter and the reverted user is significantly larger than expected. (**f**) The *AB-CA* motif occurs more commonly than expected. It is significantly faster for the English Wikipedia here but this result is neither consistent among the other languages nor robust to different measures of the shape of the distribution. The status difference between *C* and *A* is significantly smaller than expected.

---

interpretation of the *AB-CA* motif is that when a revert occurs, a third party intervenes and questions the authority of the reverter. A priori, we expected that an editor of higher status would be more likely to contest another editor's opinion. In contrast, the status effects we observe suggest that "third-party defense" happens due to daring newcomers, rather than experienced and established users.

Discussion

Large-scale collaboration by volunteers online provides much of the information we obtain and the software products we use today. The repeated interactions of these volunteers have also given rise to communities with shared identity and practice. The social interactions in these communities can in turn induce biases and subjectivities into collaborative public goods. Identifying the social interactions that play a role in these communities is thus an important first step towards understanding the biases in the content and products we consume.

In online collaboration communities, individuals may sometimes negatively evaluate or even undo other individuals' contributions. These actions may be related to the collaborative project but may also indicate negative social interactions among individuals. To better understand the latter, we here proposed to investigate the temporal and structural patterns of sequences of content-related actions.

To exemplify this approach, we analyzed six temporal motifs in the network of reverts among Wikipedia editors. We found evidence that Wikipedia editors systematically revert the same person, revert back their reverter, and come to defend a reverted editor. We did not find evidence that they "pay forward" a revert, coordinate with others to revert an editor, or revert different editors serially. In addition, we found that reciprocation of reverts tends to occur between status equals, while those who revert the reverters of others tend to be with low status. Further, high-status contributors are more likely to be involved in serial reverts. In essence, if senior Wikipedia editors revert, they may continue and revert someone else, they may get reverted by a low status contributor, or they may get reverted back if the editor they reverted is of equal standing.

Some of our findings are intuitive in the context of the structure and organization of Wikipedia. For example, vigilantism is common on Wikipedia and in particular, senior editors are always on the lookout for potential vandals or clueless newcomers. Hence, it is not surprising that serial reverts are not systematic but mainly conducted by senior editors. There are also explicit Wikipedia rules that discourage serial reverts; in particular, the "three-revert rule" prohibits editors from performing more than three reverts on the same page within 24 hours.

The context of Wikipedia also provides alternative explanations for some of our findings. For example, the result that "third-party defenders" tend to be of low status can be due to sock-puppetry, which refers to the practice of maintaining multiple accounts on Wikipedia with the aim to disrupt discussions or avoid editorial restrictions. If this were the case, senior editors would systematically have secondary accounts which they would use to revert the reverts of the edits they have done from their primary account. Such cases undoubtedly occur in our data. Nevertheless, a lot of this secretive behavior could also be happening through anonymous accounts and our analyses excluded those. Future research should investigate whether sock-puppetry indeed dictates the patterns we observed.

Still, some of our findings are unexpected. For example, although editorial wars and back-and-forth conflict have been extensively studied before, we were surprised to find that reciprocation of reverts is common except among the editors of the Japanese and Chinese editions of Wikipedia. The Japanese and Chinese cultures are known as "honor-shame cultures" so one plausible explanation is that editors in these languages avoid direct conflict in fear of ostracism[53,54]. In general, we treated most differences by language as a robustness test, as we acknowledge that multiple hypothesis testing on large amounts of data is bound to produce significant results in either direction[55]. Nevertheless, it is worthwhile to use more robust techniques to specifically investigate any cultural differences between the different language communities on Wikipedia.

Overall, our aim was to quantify the extent to which certain types of negative social interactions play a role in online collaborative communities, using the particular example of Wikipedia. We focused our analysis on the macro-level and thus, admittedly, we miss the nuanced understanding that thick ethnographic descriptions can produce[56]. Nevertheless, our findings provide the foundation and hopefully, an inspiration, for such more focused studies in the future. Our finding that certain types of negative social interactions and status considerations interfere with knowledge production on Wikipedia has practical implications both for controlling the quality of content and for maintaining editors' productive involvement on Wikipedia. Future research should use more in-depth analyses at a smaller scale to verify and qualify our findings.

Although our study revealed new knowledge about the structure and dynamics of negative interactions on Wikipedia, our findings do not automatically extend to other social systems. Still, in alignment with Keegan
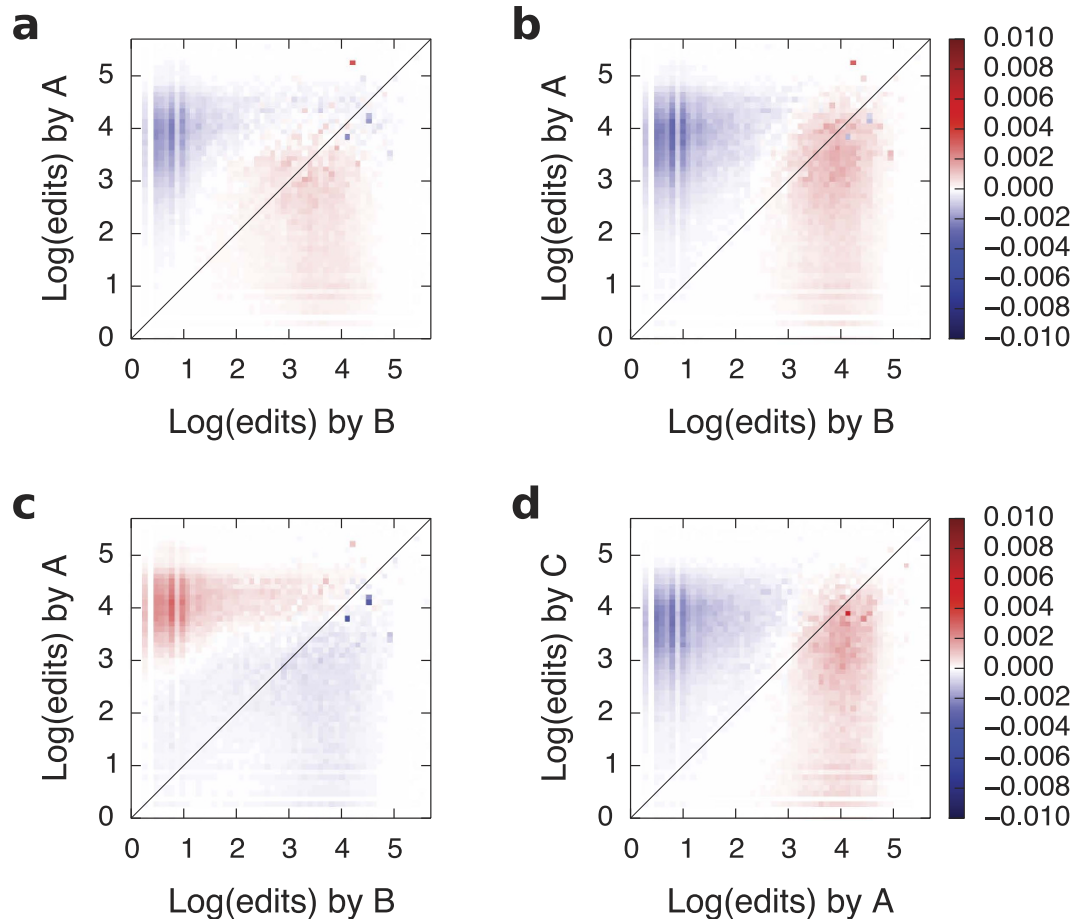
Figure 4. Difference between the observed and expected distribution of status in four of the six temporal motifs. (**a**) For the *AB-BA* motif, *A* is more likely to be of equal or lower status than *B* compared to all other reverts. (**b**) For the *AB-BC* motif, *A* is more likely to be of equal or lower status than *B*. (**c**) For the *AB-AC* motif, *A* is more likely to have high status. (**d**) For the *AB-CA* motif, *C* is more likely to be of equal or lower in status than *A*.

et al.[41], we strongly believe in the research potential of sequence and temporal motif analysis for illuminating this research problem. Thus, future research should apply our approach to open-source software projects and bulletin-board-type systems to examine the extent to which our findings generalize to other collaborative communities online.

Methods

Wikipedia data.   Our data contain who reverts whom, when, and in what article. To obtain this information, we analyzed the Wikipedia XML Dumps (https://dumps.wikimedia.org/mirrors.html) of the 13 language editions we study. To detect restored versions of an article, a hash was calculated for the complete article text following each revision and the hashes were compared between revisions. To create the network, we assumed that a link goes from the editor who restored an earlier version of the article (the "reverter") to the editor who made the revision immediately after that version (the "reverted"). In an alternative operationalization, we could have created links from the reverter to every editor who made a revision in the time between the restored version and the revert but this would have resulted in considerably denser networks.

The revert networks were then pruned by removing self-reverts and by removing anonymous users, vandals, and bots, as well as any resulting isolates. We identified anonymous users as editors who did not use a registered username to revert (the data include an IP address instead) and vandals as editors who had all their edits reverted by others. The latter rule meant that we also got rid of newcomers who became discouraged and left Wikipedia after all their initial contributions were reverted. Since we are interested in social interactions emerging from repeated activity, we do not believe that this decision affects our results. To remove the bots, we identified all accounts with bot status (https://en.wikipedia.org/wiki/Wikipedia:Bots/Status) in the Wikipedia database (as of August 6, 2015). In addition, we removed any accounts with a username containing different spelling variations of the word "bot."

Null model.   A number of sophisticated statistical methods for analyzing network data already exist but they are not adapted for large growing networks observed over a long continuous period of time. Exponential

random-graph models (ERGMs) cannot account for the timing of interactions because they express the structural properties of aggregate networks or networks observed at a single moment in time[42]. Extensions such as temporal ERGMs[57] and stochastic actor-based models[58] account for network dynamics but are nevertheless restricted to relatively small networks with a fixed set of nodes and a few "snapshots" over time.

To establish the statistical significance of our observations, we use a null model created by randomizing the underlying network. The randomization needs to preserve the daily pattern and the community structure of the network, while removing any systematic clustering in the timing of events in which an individual is involved, as we consider such clustering evidence for a social process. Hence, we do not randomize the network structure but only the timing of events. There are several possible ways to do this[18,59]. First, one can shuffle the entire time sequence of events but this destroys individual activity patterns and increases the time variation in the individual's activities, as many editors joined Wikipedia for a limited period only and the shuffle is over almost 11 years. Individual activity patterns can be preserved if the randomization instead repeatedly samples two random nodes with equal number of events and swaps the time sequences for the events. The same could be achieved by shuffling within shorter periods of time, say 24 or 168 hours.

In addition to swapping and shuffling within a node, one can swap and shuffle within dyad, only within a node's in-links, or only within a node's out-links. These methods preserve the activity patterns characteristic for dyadic exchange, individual "visibility", and individual activity, respectively. However, in this way they also restrict the baseline to these narrow scopes. For example, swapping within dyads would compare the occurrence of the *AB-AB* motif to the *AB-BA* motif but would not answer whether either of them is more common than chance.

To account for these potential caveats, we choose to use as a baseline a shuffle of the timestamps of events within individuals and within a window of 24 hours. For each link *l*, we look at the source *i* and collect other links in which *i* participated (as either source or target) up to 24 hours before or after *l*. We then swap *l*'s time with the time of one of these links selected at random; if the set of candidate links is empty, no swap occurs. We shuffle within individuals to preserve individual activity patterns and to identify each social interaction against this expected sequence of events. We execute the shuffle within a limited time window to preserve the duration of activity per individuals. The time window should be at least 24 hours to allow for interactions between editors from different time zones and with different daily routines. We found that a time window on the order of 24–240 hours produces similar results. Hence, we chose 24 hours as this is the time window we use to define the social interactions.

In short, our shuffling method does not change the structure of the network (who reverts whom), which implies that it preserves the community structure centered around articles and topics. Neither does the method change the overall sequence of timestamps, thus preserving any natural burstiness of activity due to editors being in the same time zone or due to the occurrence of news-worthy events, for example. In contrast, the method deliberately shuffles the sequence of reverts that an editor is involved in, thus removing any individual behavioral patterns.

Statistical tests.   The shuffled network provides us with an expected distribution for each interaction over the period of 24 hours. Comparing the weight of the distribution in the observed data with that in multiple shuffled realizations of the network can tell us the extent to which the interaction is more common than expected. Similarly, comparing the shape of the distribution can tell us the extent to which the interaction occurs at a faster rate than expected. To quantify the comparison, we estimate the *z*-score as follows:

$$Z = (X_D - \mu(X_R))/\sigma(X_R). \tag{1}$$

In equation (1), X_D is the relevant statistic in the data (for example, count or mean), X_R is the same statistic in the shuffled networks, μ is the mean, and σ is the standard deviation.

The statistic we use for frequency is the count of events (Supplementary Table S1). Two statistics we use for rate include the mean and skewness. Alternatively, to compare the shapes of two distributions, we also conduct a two-sample Kolmogorov-Smirnoff (KS) test between the data and one randomly chosen baseline generation. The KS test measures the maximum area deviation between the two normalized cumulative distributions. The sign of the KS statistic tells us the direction of the difference while the *p*-value tells us the likelihood of the observed distance given the null hypothesis of the distributions being the same. The problem with the KS test is that it is extremely sensitive to small deviations and thus overestimates the significance of the differences. Hence, we use the signed KS statistic to report the results but we use the mean and skewness *z*-scores to discuss their significance. Overall, the results are similar in terms of direction across the three statistics (Supplementary Table S2).

Status.   To investigate the effect of status in the interactions, we operationalize status as the base-ten logarithm of the number of edits the editor has completed by the time of the revert under question. We transformed the number of edits with the logarithm because they follow a power-law distribution (Supplementary Fig. S18). As a result, the difference in status can then also be expressed as the base-ten logarithm of the ratio of number of edits. To find evidence that status plays a role, we need to compare the status difference in the focal interaction with the status difference in any other revert. The observations are not independent across and within these two groups, however, but nested within individuals. To account for this, we execute regression analyses with standard errors adjusted for clustering within reverter and reverted (Supplementary Table S3).

References

1. S. P. Borgatti, A. Mehra, D. J. Brass & G. Labianca. Network analysis in the social sciences. *Science*. **323,** 892–895 (2009).
2. N. Eagle, A. S. Pentland & D. Lazer. Inferring friendship network structure by using mobile phone data. *Proc. Natl. Acad. Sci. USA* **106,** 15274–15278 (2009).
3. A. Barabási *et al.* Evolution of the social network of scientific collaborations. *Phys. A Stat. Mech. its Appl.* **311,** 590–614 (2002).

4. K. Rook. The negative side of social interaction: impact on psychological well-being. *J. Pers. Soc. Psychol.* **46,** 1097–1108 (1984).
5. R. Baumeister & E. Bratslavsky. Bad is stronger than good. *Rev. Gen. Psychol.* **5,** 323–370 (2001).
6. S. Cohen. Social relationships and health. *Am. Psychol.* **59,** 676–684 (2004).
7. G. F. Brenner, N. K. Norvell & M. Limacher. Supportive and problematic social interactions: A social network analysis. *Am. J. Community Psychol.* **17,** 831–836 (1989).
8. K. A. Jehn, G. B. Northcraft & M. A. Neale. Why differences make a difference: A field study of diversity, conflict, and performance in workgroups. *Adm. Sci. Q.* **44,** 741 (1999).
9. F. Heider. Attitudes and cognitive organization. *J. Psychol.* **21,** 107–112 (1946).
10. D. Cartwright & F. Harary. Structural balance: a generalization of Heider's theory. *Psychol. Rev.* **63,** 277–293 (1956).
11. J. A. Davis. Clustering and structural balance in graphs. *Hum. Relations.* **20,** 181–187 (1967).
12. J. Leskovec, D. Huttenlocher & J. Kleinberg. In *Proceedings of the 28th international conference on Human factors in computing systems-CHI '10* pp. 1361–1370 (ACM Press, New York, New York, USA, 2010).
13. J. Kunegis, A. Lommatzsch & C. Bauckhage. In *Proceedings of the 18th international conference on World wide web-WWW '09*, pp. 741–750 (ACM Press, New York, New York, USA, 2009).
14. S. Maniu, B. Cautis & T. Abdessalem. In *Databases and Social Networks on-DBSocial '11* pp. 19–24 (ACM Press, New York, New York, USA, 2011).
15. G. Facchetti, G. Iacono & C. Altafini. Computing global structural balance in large-scale signed social networks. *Proc. Natl. Acad. Sci. USA* **108,** 20953–20958 (2011).
16. S. A. Marvel, J. Kleinberg, R. D. Kleinberg & S. H. Strogatz. Continuous-time model of structural balance. *Proc. Natl. Acad. Sci. USA* **108,** 1771–1776 (2011).
17. M. Szell, R. Lambiotte & S. Thurner. Multirelational organization of large-scale social networks in an online world. *Proc. Natl. Acad. Sci. USA* **107,** 13636–13641 (2010).
18. L. Kovanen, K. Kaski, J. Kertész & J. Saramäki. Temporal motifs reveal homophily, gender-specific patterns, and group talk in call sequences. *Proc. Natl. Acad. Sci. USA* **110,** 18070–18075 (2013).
19. P. Holme & J. Saramäki. Temporal networks. *Phys. Rep.* **519,** 97–125 (2012).
20. J. Giles. Internet encyclopaedias go head to head. *Nature.* **438,** 900–901 (2005).
21. E. S. Callahan & S. C. Herring. Cultural bias in Wikipedia content on famous persons. *J. Am. Soc. Inf. Sci. Technol.* **62,** 1899–1915 (2011).
22. A. Samoilenko & T. Yasseri. The distorted mirror of Wikipedia: a quantitative analysis of Wikipedia coverage of academics. *EPJ Data Sci.* **3,** 1 (2014).
23. E. Borra *et al.* In *Proceedings of The International Symposium on Open Collaboration-OpenSym '14* pp. 1–1 (ACM Press, New York, New York, USA, 2014).
24. E. Borra *et al.* In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems-CHI '15 pp. 193–196 (ACM Press, New York, New York, USA, 2015).
25. M. Restivo & A. van de Rijt. Experimental study of informal rewards in peer production. *PLoS One.* **7,** e34358 (2012).
26. C. Danescu-Niculescu-Mizil, L. Lee, B. Pang & J. Kleinberg. In *Proceedings of the 21st international conference on World Wide Web-WWW '12* pp. 699–708 (ACM Press, New York, New York, USA, 2012).
27. A. van de Rijt, S. M. Kang, M. Restivo & A. Patil. Field experiments of success-breeds-success dynamics. *Proc. Natl. Acad. Sci. USA* **111,** 6934–6939 (2014).
28. O. Arazy, L. Yeo & O. Nov. Stay on the Wikipedia task: When task-related disagreements slip into personal and procedural conflicts. *J. Am. Soc. Inf. Sci. Technol.* **64,** 1634–1648 (2013).
29. K. Osman. In *Proceedings of the 9th International Symposium on Open Collaboration-WikiSym '13* pp. 1–6 (ACM Press, New York, New York, USA, 2013).
30. W. Ben Towne, A. Kittur, P. Kinnaird & J. Herbsleb. In *Proceedings of the 2013 conference on Computer supported cooperative work-CSCW '13* p. 1059 (ACM Press, New York, New York, USA, 2013).
31. 31. A. Kittur, B. Suh, B. A. Pendleton & E. H. Chi. In *Proceedings of the SIGCHI conference on Human factors in computing systems-CHI '07* pp. 453–462 (ACM Press, New York, New York, USA, 2007).
32. A. Halfaker, A. Kittur & J. Riedl. In *Proceedings of the 7th International Symposium on Wikis and Open Collaboration-WikiSym '11* p. 163 (ACM Press, New York, New York, USA, 2011).
33. 33. U. Brandes, P. Kenis, J. Lerner & D. van Raaij. In *Proceedings of the 18th international conference on World wide web-WWW '09* pp. 731–740 (ACM Press, New York, New York, USA, 2009).
34. R. Sumi, T. Yasseri, A. Rung, A. Kornai & J. Kertész. In *Proceedings of the ACM WebSci '11* pp. 1–3 (Koblenz, Germany, 2011).
35. T. Yasseri, R. Sumi, A. Rung, A. Kornai & J. Kertész. Dynamics of conflicts in Wikipedia. *PLoS One.* **7,** e38869 (2012).
36. Y. Gandica, F. Sampaio dos Aidos & J. Carvalho. The dynamic nature of conflict in Wikipedia. *EPL (Europhysics Lett.* **108,** 18003 (2014).
37. S. DeDeo. Conflict and computation on Wikipedia: A finite-state machine analysis of editor interactions. *arXiv Prepr. arXiv1512.04177* (2015).
38. J. Török *et al.* Opinions, conflicts, and consensus: modeling social dynamics in a collaborative environment. *Phys. Rev. Lett.* **110,** 088701 (2013).
39. G. Iñiguez, J. Török, T. Yasseri, K. Kaski & J. Kertész. Modeling social dynamics in a collaborative environment. *EPJ Data Sci.* **3,** 7 (2014).
40. D. Jurgens & T.-C. Lu. In *Proceedings of the Sixth International AAAI Conference on Weblogs and Social Media* (AAAI, 2012), pp. 162–169.
41. B. C. Keegan, S. Lev & O. Arazy. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing-CSCW '16* pp. 1063–1077 (ACM Press, New York, New York, USA, 2016).
42. G. Robins, P. Pattison, Y. Kalish & D. Lusher. An introduction to exponential random graph (p*) models for social networks. *Soc. Networks.* **29,** 173–191 (2007).
43. T. A. B. Snijders, P. E. Pattison, G. L. Robins & M. S. Handcock. New Specifications for exponential random graph models. *Sociol. Methodol.* **36,** 99–153 (2006).
44. L. Kovanen, M. Karsai, K. Kaski, J. Kertész & J. Saramäki. Temporal motifs in time-dependent networks. *J. Stat. Mech. Theory Exp.* **2011,** P11005 (2011).
45. M. Burke & R. Kraut. In *Proceedings of the ACM 2008 conference on Computer supported cooperative work-CSCW '08* pp. 27–36 (ACM Press, New York, New York, USA, 2008).
46. J. Leskovec, D. Huttenlocher & J. Kleinberg. Governance in social media: A case study of the Wikipedia promotion process. *ICWSM* (2010).
47. L. Festinger. A theory of social comparison processes. *Hum. Relations.* **7,** 117–140 (1954).
48. G. J. Kilduff, H. A. Elfenbein & B. M. Staw. The psychology of rivalry: A relationally dependent analysis of competition. *Acad. Manag. J.* **53,** 943–969 (2010).
49. J. M. Hoobler & D. J. Brass. Abusive supervision and family undermining as displaced aggression. *J. Appl. Psychol.* **91,** 1125–1133 (2006).

50. R. M. Sapolsky. Culture in animals: The case of a non-human primate culture of low aggression and high affiliation. *Soc. Forces*. **85,** 217–233 (2006).
51. T. Yasseri & J. Kertész. Value production in a collaborative environment. *J. Stat. Phys*. **151,** 414–439 (2013).
52. M. B. Lieberman & S. Asaba. Why do firms imitate each other? *Acad. Manag. Rev.* **31,** 366–385 (2006).
53. J. Y. Kim & S. H. Nam. The concept and dynamics of face: Implications for organizational behavior in Asia. *Organ. Sci.* **9,** 522–534 (1998).
54. M. W. Pye & L. W. Pye. *Asian Power and Politics: The Cultural Dimensions of Authority* (Harvard University Press, 2009).
55. D. Ruths & J. Pfeffer. Social sciences. Social media for large studies of behavior. *Science*. **346,** 1063–1064 (2014).
56. D. Jemielniak. *Common Knowledge?: An Ethnography of Wikipedia* (Stanford University Press, Stanford, CA, 2014).
57. P. N. Krivitsky & M. S. Handcock. A separable model for dynamic networks. *J. R. Stat. Soc. Series B. Stat. Methodol.* **76,** 29–46 (2014).
58. T. A. B. Snijders, G. G. van de Bunt & C. E. G. Steglich. Introduction to stochastic actor-based models for network dynamics. *Soc. Networks*. **32,** 44–60 (2010).
59. M. Karsai *et al.* Small but slow world: How network topology and burstiness slow down spreading. *Phys. Rev. E.* **83,** 025102 (2011).

Acknowledgements

Author Contributions

M.T. and T.Y. designed and performed research. M.T. analyzed data. M.T., R.G.-G. and T.Y. wrote and reviewed the manuscript.

Additional Information

Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Tsvetkova, M. *et al.* Dynamics of Disagreement: Large-Scale Temporal Network Analysis Reveals Negative Interactions in Online Collaboration. *Sci. Rep.* **6**, 36333; doi: 10.1038/srep36333 (2016).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.