

The Network Structure of Successful Collaboration in Wikipedia

Jürgen Lerner
University of Konstanz
juergen.lerner@uni-konstanz.de

Alessandro Lomi
University of Lugano
alessandro.lomi@gess.ethz.ch

Abstract

Wikipedia is one of the largest and most successful examples of decentralized peer-production systems currently in existence. Yet, the quality of Wikipedia articles varies widely with articles considered of encyclopedic quality (called featured articles) representing less than 0.1 percent of all articles. In this paper, we examine how article quality varies as a function of the network mechanisms that control the interaction among contributors. More specifically, we compare the network mechanisms underlying the production of the complete set of featured articles, with the network mechanisms of a contrasting sample of comparable non-featured articles in the English-language edition of Wikipedia. Estimates of relational event models suggest that contributors to featured articles display greater deference toward the reputation of their team members. Contributors to featured articles also display a weaker tendency to follow the behavioral norms predicted by the theory of structural balance, and hence a weaker tendency toward polarization.

1. Introduction

The online encyclopedia Wikipedia is the largest general reference work available in the Internet and one of the most visited websites in existence. Wikipedia is also one of the success stories of self-organizing online peer-production where communities of volunteers actively sustain private costs to produce freely available public goods [1]. Despite its global popularity and success, the quality of articles in Wikipedia is highly heterogeneous [2, 3, 4, 5]. Featured articles, the set of articles considered of the highest (“encyclopedic”) quality, are less than 0.1 percent of all articles in Wikipedia. Why are high quality articles in Wikipedia so few? Why is the quality of Wikipedia articles so variable? Addressing these questions would contribute to improve our understanding of how to control

and stabilize quality in decentralized peer-production systems – a problem whose importance goes well beyond the specific case of Wikipedia that we examine in this paper.

In the near-absence of centralized, authority-based organizational coordination and control routines, it has been argued that emergent networks resulting from task-oriented interaction might serve the purpose of coordinating and controlling the execution of team work [6]. If this claim is correct, then differences in the structure of these collaboration networks should be reflected in observable differences in the quality of the articles in Wikipedia. In this paper, we explore this conjecture by comparing the structure of the *edit network* [7] of featured articles with that of a control sample of non-featured articles that have similar distributions with respect to basic characteristics.

The structure of the edit network of a Wikipedia article emerges from sequences of time-stamped, signed and weighted relational events [6]. A negative *undo event* results from edits in which a user deletes (at least part of) the contributions of another user. A positive *redo event* results from edits in which a user restores (at least part of) the contributions of another user that have been previously deleted. Undo and redo events are weighted by the number of words deleted or restored. These decisions to undo or redo contributions of other users are of utmost importance since they ultimately determine the article’s content and quality. Previous work has shown that these decisions can be partially explained by the embedding of users in the network of past events [6]. In this paper we develop and test two hypotheses about how the network structure of collaboration underlying the production of higher-quality articles may differ from that underlying the production of lower-quality articles.

1.1. Background and hypotheses

The theory of structural balance [8, 9] stipulates that a signed network (that is, a network with positive

and negative ties) is balanced if all of its cycles have an even number of negative ties. The theory assumes that individuals will have a preference for balanced structures. There is a conceptual link between polarization [10], balance [8, 9], and cognitive dissonance [11]. The structure theorem of Cartwright and Harary [9] demonstrates that a signed network is balanced if and only if its nodes partition into two classes with only positive ties within classes and only negative ties between classes. Thus, a signed network is maximally polarized if, from the point of view of each individual, it decomposes into “friends” and “enemies,” such that friends are positively linked to friends, friends are negatively linked to enemies, and enemies are positively linked to enemies. Such a network structure reduces cognitive dissonance [11] and is associated with a stable cognitive equilibrium state [8].

Our claim that a such a balanced structure might impair team productivity is based on the observation that polarization has undesirable consequences for communities. For instance, polarization has been found to lead to social segregation, biased assimilation of new information, exposure to less diverse points of view, and to cause an even higher degree of polarization [12, 13]. Similarly, it has been shown that the human tendency to reduce cognitive dissonance may have negative economic consequences [14]. Adherence to the rules predicted by balance theory also tends to perpetuate political conflict [15]. The general conclusion that we draw from this extensive literature is that if a social system is balanced, it is both stable and sub-optimal.

More specifically for the case of Wikipedia, strong opinion polarization within the team of contributors to an article, may push users to evaluate the contributions of others not only in terms of their quality, but also in the light of ingrained ideological orientations. Users might be willing to delete even high-quality content if it expresses opposing views, or simply if it originates from members of the “out-group.” Such a balance-driven editing behavior would undermine the production of a high-quality article. *Therefore, we hypothesize that editing behavior underlying the production of high-quality (featured) articles will display a weaker tendency to conform to the predictions of balance theory, than editing behavior underlying the production of non-featured articles.*

Important as it may be, polarization is not the only factor driving the quality of Wikipedia articles. At least as important is the quality of individual contributions – regardless of the agreement or disagreement they might generate. An alternative hypothesis could be that editing decisions based on the quality of content will be more likely to produce high quality collective outcomes.

Since it is hard to assess the quality of individual contributions, we use user reputation [16, 17] as a proxy for quality. A user has high reputation, if it is unlikely that his or her contributions are questioned and hence deleted by others. *We hypothesize that teams producing featured articles will display more deference towards (and hence more reluctance to undo) the contributions of highly reputable contributors.*

The question whether polarization (“us vs. them”) or reputation drives text-editing decisions relates closely to the distinction between balance theory and status theory analyzed by [18] on a network of votes for or against Wikipedia adminship. While [18] found more support for status theory, we want to relate article quality with these hypothetical explanations for negative ties.

The analysis of conflict, dispute, disagreement, or controversy among Wikipedia editors is not new [19, 20, 21, 22, 23, 24, 25]. In this paper, however, we are not analyzing the impact of the *presence* of, or the amount of, conflict in a team of Wikipedia contributors, but rather the effect of the *structure* of conflict. The relation between network structure and quality of Wikipedia articles is also not examined here for the first time. For example, in a paper relating network structure to the quality of Wikipedia articles, [26] analyze edit motifs in the bipartite network connecting users to articles they contribute to write. The network we analyze in this paper is different, however, as it links users through positive and negative events resulting from text modifications.

2. Wikipedia articles

The data that we analyze in this paper come from the public database dump (<https://dumps.wikimedia.org/>) of the English-language edition of Wikipedia, dated January 1st, 2018. There are about 5.5 million *articles* (that is, Wikipedia pages from Namespace 0 that are not redirects) in this data. Articles in this set are very heterogeneous with respect to their quality, but also with respect to basic variables such as size, age, or number of edits. In fact, the average article is very short and has received a very small number of edits. Since it is meaningless to compare a mature high-quality article with a newly created, or mostly neglected article consisting of only a few sentences, we take special care in constructing the sample of articles that we analyze. We define next the set of high-quality articles used in our analysis and in Section 2.2 we describe the construction of a comparison sample of articles that are not perceived as being of high quality – but that are similar to the high-quality articles with respect to several basic variables.

2.1. High-quality (featured) articles

Wikipedia has a well-established internal, community-based quality evaluation system for its articles. Most articles are assigned to one (or multiple) quality levels: stub, start, C-class, B-class, good article (GA), A-class, or featured article (FA). Wikipedia’s quality assessment has been compared to external ratings on a sample of articles by [2], who generally found satisfactory agreement, and has often been used in academic research [2, 26, 4, 27, 28].

In this paper we analyze all featured articles (FA) in existence as of January 1st, 2018. This set of 5,164 featured articles constitutes half of our article sample. The non-featured articles that we selected for case control completes the full sample of articles that we analyze.

2.2. Constructing a comparison set of non-featured articles

Comparing featured articles with arbitrary (or random) non-featured articles would be futile since a typical featured article has by several orders of magnitude more edits, more contributors, and a longer text than a typical non-featured article. In fact, a randomly drawn article has only a tiny chance of being featured – not because the network among its collaborators has the “wrong” structure but simply because it is much too short to qualify as a featured article (compare [29]).

In order to uncover the distinctive features of the collaboration structure of high-quality articles, we compare all featured articles with a control sample of non-featured articles having similar distributions with respect to several variables that are predictive of the probability to be featured. Such an approach has been advocated, for instance, by [5] who compare “coordination” in Wikipedia’s featured articles with a comparison set of non-featured articles that have similar distributions of edits. In this paper we choose the comparison set of non-featured articles dependent on many more variables.

Our approach to constructing a control sample of non-featured articles may be summarized as follows. We first specify a basic model for the probability that a Wikipedia article is featured and fit this model on all 5.5 million articles. Explanatory variables in this basic model include article characteristics that are reliable predictors of article quality, but that do not encode the structure of the collaboration network. We then draw about 5,000 non-featured articles such that the distribution of the predicted FA-probability on these sampled non-featured articles is similar to the predicted

Table 1. Logistic regression for the probability of being featured, using basic explanatory variables, estimated on all Wikipedia articles.

(Intercept)	−10.938 (0.055)*
log1p.length	2.471 (0.034)*
age	0.905 (0.025)*
log1p.#edits	1.585 (0.046)*
log1p.#reverts	0.971 (0.023)*
log1p.teamsize	−2.583 (0.045)*
log1p.#wiki.links	−0.769 (0.031)*
log1p.#external.refs	−0.139 (0.017)*
log1p.#lang.links	0.318 (0.018)*
#level.1.sections	−0.305 (0.015)*
#level.2.sections	−0.444 (0.011)*
log1p.#images	0.160 (0.015)*
log1p.#templates	0.669 (0.031)*
#characters.per.word	−0.667 (0.024)*
#words.per.sentence	−0.226 (0.042)*
#categories	0.054 (0.006)*
log1p.avg.cat.size	−0.084 (0.024)*
granularity	0.475 (0.030)*
AIC	44,373
BIC	44,617
Num. obs.	5,521,800

* $p < 0.001$

FA-probability of the featured articles. Thus, judging from the basic model, non-featured articles included in the control sample have comparable odds of being included in the set of featured articles – except for the obvious fact that they are not actually included in it. We then analyze differences in the network structure of featured and sampled non-featured articles in order to detect characteristics that might explain the difference in quality. Details of this approach are discussed next.

As basic explanatory variables for article quality we take the articles’ length (number of bytes); age (time since the first edit); number of edits; team size (number of unique contributors); numbers of reverts, links within Wikipedia, external references, sections at level one and two, images, templates, categories, and links to articles in other language editions of Wikipedia; the average number of characters per word and the average number of words per sentence (compare [4]); the average size of categories; and a measure for the article granularity (compare [28]). Variables that have a skewed distribution are transformed by the mapping $x \mapsto \log(1 + x)$ and all variables are normalized by subtracting their mean and dividing by their standard deviation.

We use these variables to estimate a logit model for the probability that given articles are featured.

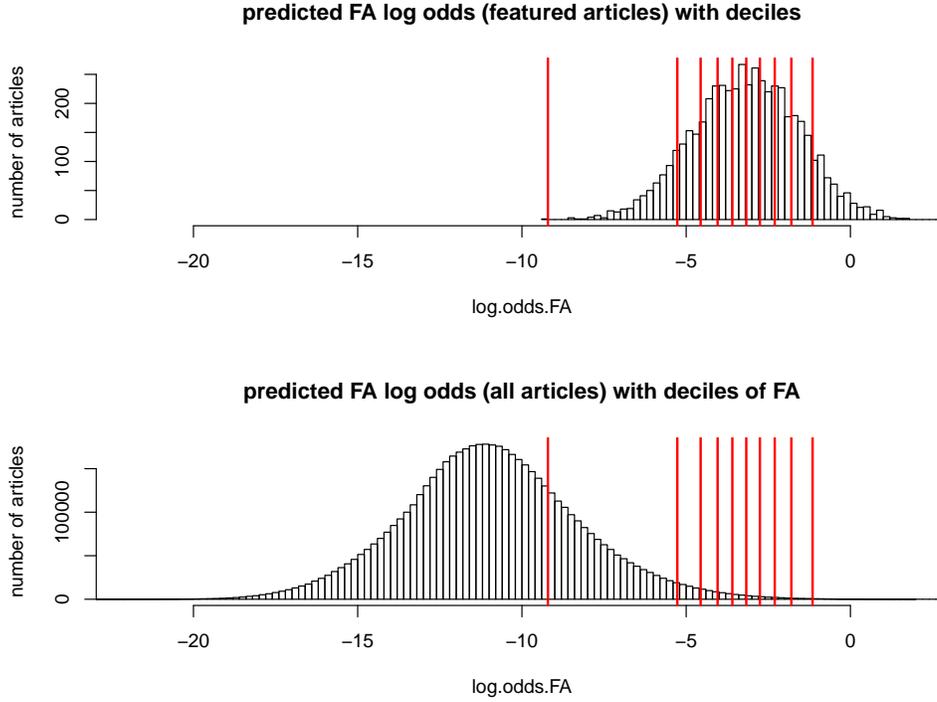


Figure 1. Histogram of log-odds for being featured as predicted by the model from Table 1. **Top:** predicted log-odds for all 5,164 featured articles. The vertical lines show the deciles $(q_i)_{i=0,\dots,9}$ of featured articles. **Bottom:** predicted log-odds for all 5.5 million articles. The vertical lines also show the deciles computed from the set of featured articles (not from all articles). For each $i = 0, \dots, 9$, we sample about 516 non-featured articles whose predicted log-odds are in the interval $[q_i, q_{i+1})$.

Estimated parameters, obtained from fitting the model to all 5.5 million articles, are reported in Table 1. Many of the findings from this model are well-known or are consistent with intuition.

The model shown in Table 1 predicts for all articles the probabilities to be featured whose log-odds are displayed in Fig. 1. The vertical lines show the deciles $(q_i)_{i=1,\dots,9}$ of featured articles that is, the values q_i such that $(i \cdot 10)\%$ of all featured articles have predicted log-odds below q_i . For ease of notation, we define q_0 to be the minimum predicted log-odds over all featured articles and q_{10} to be the maximum plus one.

To construct a case control sample of non-featured articles, we first discard all articles whose predicted log-odds are below q_0 . For $i = 0, \dots, 9$ let A_i denote the set of all non-featured articles a whose predicted log-odds satisfy

$$q_i \leq \log.odds(a) < q_{i+1} .$$

For each article $a \in A_i$ we decide uniformly and independently at random with probability $p = 516.4/|A_i|$ whether to include a in our comparison

set. Thus, in expectation we draw 516.4 non-featured articles (that is, 10% of the number of all featured articles) from each of the sets A_i . Judging from the model in Table 1, these sampled non-featured articles have similar predicted probabilities to be of high quality as the featured articles. Empirically the sampled non-featured articles also have similar distributions than the featured articles in the characteristics used to fit that model. For instance, they tend to be much longer and to have many more revisions than the average over all 5.5 million articles. As it could be expected, the quality distribution of the sampled non-featured articles is centered at the higher quality grades below FA (that is, B-class, good article, and A-class), while the quality distribution of all 5.5 million articles has a pyramidal shape with stub articles being the most frequent, followed by start and C-class articles.

3. Modeling Wikipedia edit networks

For each Wikipedia article in our sample (featured or not) we compute an *edit network* [7] that is

composed of time-stamped, signed, and weighted relational events resulting from text modifications. Undoing contributions of others is interpreted as a negative interaction among users and re-doing others’ contributions is interpreted as a positive interaction. Analyzing fine-grained text-change events obtained by comparing successive revisions of the same article is a well established strategy in Wikipedia research and yields more detailed information than considering just complete reverts that restore exactly a previous version of the text [16, 7, 17, 30, 31, 6].

We model the propensity of specific users to undo contributions of specific other users by variants of relational event models that have been proposed in [6]. Relational event models [32] are a family of statistical models to analyze networks of social interaction defined in terms of dyadic, time-stamped events observed in continuous time, such as persons sending emails to others – or Wikipedia users undoing contributions of other users. Relational event data contain fine-grained time information that is particularly relevant in online settings where interaction is routinely logged by servers. Models for *signed* relational events have been proposed to analyze positive and negative interaction in applications ranging from literature reviews and political discourse over to international relations [33, 34, 35, 36, 37].

The model for Wikipedia edit networks that we adopt in this paper assumes that for each pair of users (A, B) contributing to a given article and for each point in time t there is a latent probability $prob.undo_t(A, B)$ explaining how likely A is going to undo contributions of B at time t . What we can observe instead is the fraction of B ’s edits undone by A at time t , compared to the total amount of B ’s edits that could potentially be undone at t . For instance, at time t , there might be 100 words contributed by B at some time before t that are still in the article’s text at t . User A – the one who is uploading a new revision at time t – might decide to delete 25 out of these 100 words, giving an observed undo ratio of 0.25. The undo model tries to estimate the latent probabilities $prob.undo_t(A, B)$, as a function of explanatory variables introduced below, such that the estimated undo probabilities are as close as possible to the observed undo ratios. The predicted probabilities $prob.undo_t(A, B)$ are estimated by logistic regression, where the explanatory variables can be functions of sequences of past events, but can also encode characteristics of A or B , see Sect. 3.1 and Fig. 2. If the estimated parameter α_i associated with a particular explanatory variable x_i is positive, then a high value of x_i for the dyad (A, B) implies an increase in the predicted probability that A will undo edits previously

contributed by B , and hence a more negative assessment of B ’s edits by A . A negative parameter implies a decreased undo probability and, thus, a more positive assessment by A of the text contributed by B .

When constructing the edit network we did not exclude bots (i. e., software scripts that perform routine tasks [38]). While it is questionable whether software adheres to the same social norms as humans, bots are nevertheless developed and deployed by humans. Two contentious factions of bots, or a group of bots being in opposition with a group of human users, could still indicate polarization in the creation of an article. Future work has to test against the conjecture that a different ratio of non-human users, or more generally a different distribution of user roles, could explain observed differences in the structure of edit networks.

3.1. Explanatory variables in the undo model

Network effects. The actual explanatory variables that encode specific network effects are illustrated in Fig. 2. The configurations in the top row in Fig. 2 illustrate effects in which the estimated undo probability from A to B depends on past undo or redo events on the same dyad (repetition effects) or on the reverse dyad (reciprocation effects). Additional degree-based effects, illustrated in the second and third row in Fig. 2, are included to control for activity and popularity effects.

Of particular interest for the argument that we test in this paper are the triadic effects illustrated in the bottom row in Fig. 2. These explanatory variables can model how the undo probability from A to B depends on indirect ties connecting A and B via third parties. For these variables we consider undirected past undo or redo events, that is, for two users C_1 and C_2 the symmetric undo weight connecting C_1 with C_2 is defined by aggregating past undo events from C_1 to C_2 and from C_2 to C_1 (similarly for the redo weight).

Concretely, the variable *friend.of.friend* encodes to what extent A is connected by past redo events with a third user C who, in turn, is connected by past redo events with B . Structural balance theory [8, 9] predicts that “friends of friends are friends” so that if the *friend.of.friend* variable takes a high value on the dyad (A, B) we would expect a rather positive assessment of B by A , that is a low undo probability. In the estimated models this would be reflected in a negative parameter associated with *friend.of.friend*.

The variable *friend.of.enemy* encodes to what extent A is connected by past undo events with a third user C who, in turn, is connected by past redo events with B . Structural balance theory predicts that “friends of enemies are enemies” so that we expect a positive

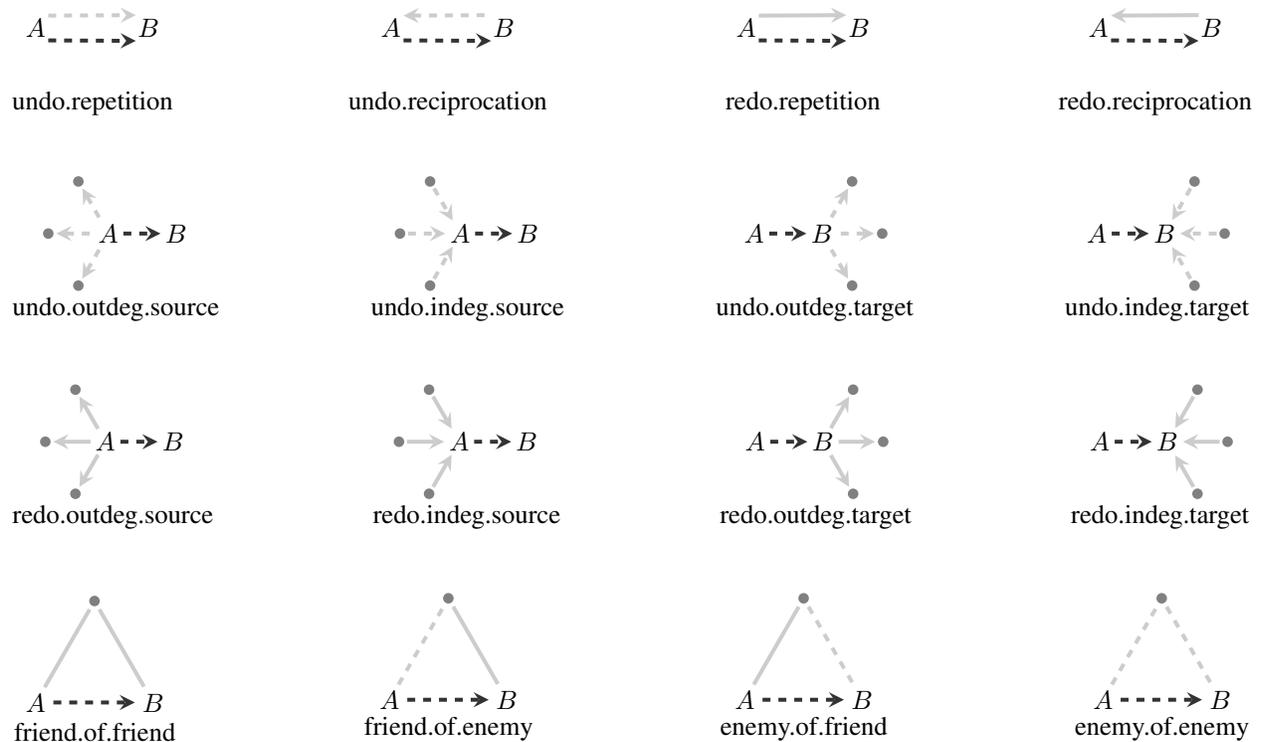


Figure 2. Local configurations of past undo events (dashed, light gray edges) or past redo events (straight, light gray edges) explaining future undo events from A to B (dashed, dark gray edges). Edges with an arrowhead aggregate only events pointing in that direction; edges without an arrowhead (used in the triadic configurations in the bottom row) aggregate events between the two incident nodes pointing in either direction.

parameter associated with *friend.of.enemy* in the undo model (revealing an increased undo probability towards the friend of an enemy). Similarly, a positive parameter is predicted for enemies of friends which should also be enemies, according to balance theory.

Finally, the variable *enemy.of.enemy* encodes to what extent A is connected by past undo events with a third user C who, in turn, is connected by past undo events with B . Structural balance theory predicts that “enemies of enemies are friends” so that we expect a negative parameter associated with *enemy.of.enemy* in the undo model (revealing a reluctance to undo contributions of enemies of enemies).

Other effects. In addition, we include as explanatory variables the number of users (nodes) in the network at the time of the event and two variables dependent on the *reputation* [16, 17] of the source user and the target user of dyadic events. The reputation of a Wikipedia user intuitively encodes the ratio of the user’s contributions that have been preserved in the past. Thus, contributions of a user with the highest reputation (equal to 1.0) have never been undone in

the past and all contributions of a user with the lowest reputation (equal to 0.0) have been undone in the very next revision. The longer the contributions of a user survive before deletion, the more her reputation approaches one. (Future work might consider capturing reputation by other user characteristics.) By common sense the reputation of a user B should decrease the probability that B ’s edits are undone, so that we expect a negative parameter associated with *reputation.of.target* in the undo model.

Distinguishing effects in featured and non-featured articles. Finally, in order to analyze whether collaboration networks on featured articles display systematically different patterns of interaction we include a binary explanatory variable that is one if the article on which the event takes place is featured and zero else. The FA indicator variable is interacted with all other explanatory variables to estimate the difference in network effects between featured and non-featured articles. Based on our previous arguments we hypothesize that it is positive for article quality if users display deference for the reputation of other

users. Thus we expect that in featured articles users will show a stronger reluctance to undo contributions of other users with high reputation. That is, we expect the parameter associated with *reputation.of.target* to be lower (more negative) on featured articles. Likewise, we hypothesize that adherence to behavioral norms predicted by structural balance theory is negative for article quality as it might indicate polarization of the team of contributors. Thus, we expect that on featured articles the parameters of the *friend.of.friend* and the *enemy.of.enemy* variables are higher than on non-featured articles (users are less reluctant to undo contributions of friends of friends and of enemies of enemies) and that the parameters of *friend.of.enemy* and *enemy.of.friend* are lower (users show a weaker tendency to fight friends of enemies or enemies of friends).

Preprocessing variables and sampling observations.

Due to skewed distributions we transform all explanatory variables, except the binary featured article indicator (FA), by the mapping $x \mapsto \log(1 + x)$ and then subtract their mean and divide by their standard deviation. These normalizations make parameter sizes more easily interpretable and comparable. To reduce runtime and memory consumption when computing explanatory variables and estimating model parameters, we apply two sampling strategies. The first deals with the extreme sparsity of edit networks. In fact when a particular user produces the next revision of an article she could potentially undo contributions of hundreds of other users but typically decides to undo only few of them (compare [6]). To deal with such situations, case-control sampling [39] has been suggested which proceeds by including all events (dubbed “failures” in [39]) but only a certain number of the non-events. Sampling from the non-events in the context of relational event models has been proposed for instance in [32, 40]. In our study we include for each observed undo event five dyads on which an undo event could have happened at that point in time, but did not. By this procedure we obtain more than 78 million observations, which is still too much to be kept in main memory during parameter estimation. To further reduce the number of observations we sample uniformly at random with a probability of 0.1 yielding 7,891,113 dyadic observations on which we estimate the undo models.

4. Results

Table 2 reports estimated parameters of the models explaining dyadic undo probabilities. We will focus our

discussion on the variables related to our hypotheses and discuss the “basic undo model” (left column) first. The finding on the effect of the *reputation.of.target* variable is consistent with the hypothesis that users, in general, show deference for the reputation of others. The negative parameter associated with this variable indicates that contributions of users with high reputation have a lower probability to be made undone. The findings on triadic network effects support – but not without exception – the predictions of balance theory. The negative parameter of *friend.of.friend* implies a decreased probability to undo the contributions of friends of friends. Thus, friends of friends tend to be seen as friends. The finding contradicting the predictions of balance theory is the negative parameter associated with *friend.of.enemy*, implying that users have a reluctance to undo contributions of the friends of their enemies. In contrast, the positive parameter associated with *enemy.of.friend* supports balance theory predicting that these are enemies (whose contributions should thus be made undone). Also in support of balance theory is the negative parameter associated with *enemy.of.enemy* indicating that users are reluctant to undo contributions of users that have the same enemies.

Turning to the model reported in the second column of Table 2 (interacting all explanatory variables with the featured article indicator), we can detect systematic differences between featured and non-featured articles. The hypothesis that users contributing to featured articles display more deference for the reputation of other contributors gets supported by the negative parameter associated to the interaction effect *reputation.of.target:FA*. This parameter implies that, when working on featured articles, users tend to be even more reluctant (than on non-featured articles) to undo contributions of other users with high reputation.

With respect to the difference on the triadic network effects on featured and non-featured articles, we hypothesized that users working on featured articles will have a weaker tendency to act according to the rules predicted by balance theory. Three of the four triadic effects support this hypothesis, the exception being the parameter for *friend.of.friend*. We estimated negative parameters associated with the interaction effects *friend.of.enemy:FA* and *enemy.of.friend:FA*. This means that users working on featured articles are more reluctant to undo contributions of the friends of their enemies and of the enemies of their friends, than users working on non-featured articles. Since balance theory predicts that “friends of enemies, and enemies of friends are enemies,” we thus find a weaker agreement with balance theory in the creation of featured articles. Going in the same direction, the parameter of the interaction

Table 2. Logistic regression for dyadic undo probabilities. Results related to our hypotheses are in bold.

	basic undo model	interact FA indicator
(Intercept)	-1.6977 (0.0003)***	-1.7240 (0.0004)***
number.of.users	0.2386 (0.0001)***	0.2684 (0.0002)***
undo.repetition	0.0906 (0.0001)***	0.1047 (0.0002)***
undo.reciprocation	0.0822 (0.0001)***	0.0793 (0.0001)***
redo.repetition	-0.0228 (0.0001)***	-0.0159 (0.0002)***
redo.reciprocation	-0.1607 (0.0001)***	-0.1378 (0.0002)***
undo.outdegree.source	1.1462 (0.0009)***	1.1331 (0.0013)***
undo.indegree.source	-0.5343 (0.0007)***	-0.4611 (0.0010)***
undo.outdegree.target	0.1171 (0.0003)***	0.0546 (0.0004)***
undo.indegree.target	0.3251 (0.0002)***	0.3289 (0.0003)***
redo.outdegree.source	0.0893 (0.0007)***	-0.0025 (0.0010)*
redo.indegree.source	0.0670 (0.0003)***	0.0320 (0.0005)***
redo.outdegree.target	-0.0994 (0.0003)***	-0.0538 (0.0004)***
redo.indegree.target	-0.2520 (0.0002)***	-0.2916 (0.0004)***
friend.of.friend	-0.4286 (0.0003)***	-0.4197 (0.0004)***
friend.of.enemy	-0.1694 (0.0003)***	-0.1027 (0.0005)***
enemy.of.friend	0.4891 (0.0003)***	0.5320 (0.0004)***
enemy.of.enemy	-0.0795 (0.0003)***	-0.1341 (0.0005)***
reputation.of.source	-0.4369 (0.0003)***	-0.3993 (0.0004)***
reputation.of.target	-0.8270 (0.0001)***	-0.7973 (0.0001)***
FA		0.0581 (0.0005)***
number.of.users:FA		-0.0325 (0.0003)***
undo.repetition:FA		-0.0252 (0.0003)***
undo.reciprocation:FA		0.0011 (0.0002)***
redo.repetition:FA		-0.0208 (0.0003)***
redo.reciprocation:FA		-0.0524 (0.0002)***
undo.outdegree.source:FA		-0.0015 (0.0018)
undo.indegree.source:FA		-0.1198 (0.0014)***
undo.outdegree.target:FA		0.1424 (0.0006)***
undo.indegree.target:FA		-0.0104 (0.0005)***
redo.outdegree.source:FA		0.1751 (0.0014)***
redo.indegree.source:FA		0.0588 (0.0007)***
redo.outdegree.target:FA		-0.1210 (0.0006)***
redo.indegree.target:FA		0.0621 (0.0005)***
friend.of.friend:FA		-0.0061 (0.0005)***
friend.of.enemy:FA		-0.1106 (0.0006)***
enemy.of.friend:FA		-0.0813 (0.0005)***
enemy.of.enemy:FA		0.0986 (0.0006)***
reputation.of.source:FA		-0.0668 (0.0005)***
reputation.of.target:FA		-0.0536 (0.0002)***
AIC	324,301,745	323,457,055
BIC	324,302,022	323,457,610
Num. obs.	7,891,113	7,891,113

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

effect *enemy.of.enemy:FA* is positive, implying that users working on featured articles are more willing to undo contributions of enemies of their enemies, than on non-featured articles. Since balance theory predicts that “enemies of enemies are friends,” we therefore find again a weaker agreement with balance theory on featured articles. Contradicting this hypothesis, the parameter of the interaction effect *friend.of.friend:FA* is negative meaning that users working on featured articles are even more reluctant to undo contributions of friends of their friends – being in even stronger agreement with balance theory. We note that these findings support the claim of Saperstein that “*it is hard to see anything wrong with rule 1* [15, p.289]” (referring to the rule “friends of friends are friends”), but that the other three rules of balance theory are potentially problematic.

5. Discussion and conclusions

Articles in Wikipedia are produced by dispersed communities of volunteers in the near absence of hierarchical organizational structures and control routines. Our main objective in this paper was to examine differences in the network mechanisms behind observed variation in the quality of Wikipedia articles. Our starting point was the claim that networks emerging from task-oriented interaction among participants provide a governance structure that facilitates coordination and control. We exploited the social transparency properties typical of on-line peer productions to examine the link between network mechanisms and the quality of Wikipedia articles. Our analysis was guided by two general ideas. The first was that quality of the articles decreases if interaction among contributors is shaped by balance processes that induce and sustain polarization. The second idea was that the quality of the articles is likely to benefit from a system that recognizes differences in reputation among participants.

The results of our empirical study suggest that the structure of collaboration networks might have a significant impact on the ultimate quality of articles that are otherwise similar in terms of their basic editorial features. More specifically, we have found that teams involved in the production of featured articles display more deference for the reputation of their members, i. e., they are more reluctant to question the opinions of highly reputable contributors. We have also found that teams producing featured articles display weaker adherence to the behavioral norms predicted by balance theory. Because structural balance effects are known to produce polarized networks partitioned into mutually contentious factions, this result could point to a “prize

of polarization” in that polarized teams produce output of lower quality.

5.1. Limitations and future work

The empirical analysis in this paper follows the approach of a case-control study in which instances (i. e., articles) are selected based on the outcome variable (i. e., article quality). Then, we assess differences in the explanatory variable (i. e., network structure) between instances with different outcomes. As such, our analysis cannot establish a causal effect nor does it provide an indicator of effect size. Future work that attempts to overcome these drawbacks might proceed by specifying a model for the evolution of article quality in which network structure is included as an explanatory variable. Another possibility for future work is applying a more sophisticated model in which the effect of the adherence to specific social norms on article quality might depend on the type or role of users. By doing so we might filter out systematically different behavior of, for instance, bots or users that perform editorial routine tasks. Last but not least, future work could attempt to shed light on the question to what extent users take into account information about past user interaction. While this information is indeed available to everyone with access to a Web browser, it is not obvious to what degree it influences users’ editing decisions.

Acknowledgments

This work has been supported by Deutsche Forschungsgemeinschaft (DFG Grant Nr. LE 2237/2-1) and Swiss National Science Foundation (FNS Project Nr. 100018_150126).

References

- [1] E. v. Hippel and G. v. Krogh, “Open source software and the “private-collective” innovation model: Issues for organization science,” *Organization science*, vol. 14, no. 2, pp. 209–223, 2003.
- [2] A. Kittur and R. E. Kraut, “Harnessing the wisdom of crowds in Wikipedia: quality through coordination,” in *Proc. 2008 ACM conf. Computer Supported Cooperative Work*, pp. 37–46, ACM, 2008.
- [3] O. Arazy, O. Nov, R. Patterson, and L. Yeo, “Information quality in Wikipedia: The effects of group composition and task conflict,” *Journal of Management Information Systems*, vol. 27, no. 4, pp. 71–98, 2011.
- [4] S. Ransbotham and G. C. Kane, “Membership turnover and collaboration success in online communities: Explaining rises and falls from grace in Wikipedia,” *MIS Quarterly*, vol. 35, no. 3, pp. 613–627, 2011.
- [5] D. M. Romero, D. Huttenlocher, and J. M. Kleinberg, “Coordination and efficiency in decentralized collaboration,” in *Proc. 9th intl. AAAI conf. Web and Social Media*, pp. 367–376, 2015.

- [6] J. Lerner and A. Lomi, "The Third Man: Hierarchy formation in Wikipedia," *Applied Network Science*, vol. 2, no. 1, p. 24, 2017.
- [7] U. Brandes, P. Kenis, J. Lerner, and D. van Raaij, "Network analysis of collaboration structure in Wikipedia," in *Proc. 18th Intl. Conf. WWW*, pp. 731–740, ACM, 2009.
- [8] F. Heider, "Attitudes and cognitive organization," *The Journal of Psychology*, vol. 21, pp. 107–112, 1946.
- [9] D. Cartwright and F. Harary, "Structural balance: A generalization of Heider's theory," *The Psychological Review*, vol. 63, no. 5, pp. 277–293, 1956.
- [10] J.-M. Esteban and D. Ray, "On the measurement of polarization," *Econometrica*, vol. 62, no. 4, pp. 819–851, 1994.
- [11] L. Festinger, *A Theory of Cognitive Dissonance*, vol. 2. Stanford University Press, 1962.
- [12] C. G. Lord, L. Ross, and M. R. Lepper, "Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence," *Journal of personality and social psychology*, vol. 37, no. 11, p. 2098, 1979.
- [13] N. E. Friedkin, A. V. Proskurnikov, R. Tempo, and S. E. Parsegov, "Network science on belief system dynamics under logic constraints," *Science*, vol. 354, no. 6310, pp. 321–326, 2016.
- [14] G. A. Akerlof and W. T. Dickens, "The economic consequences of cognitive dissonance," *The American economic review*, vol. 72, no. 3, pp. 307–319, 1982.
- [15] A. M. Saperstein, "'The enemy of my enemy is my friend' is the enemy: Dealing with the war-provoking rules of intent," *Conflict Management and Peace Science*, vol. 21, no. 4, pp. 287–296, 2004.
- [16] B. T. Adler and L. de Alfaro, "A content-driven reputation system for the Wikipedia," in *Proc. 16th Intl. Conf. WWW*, pp. 261–270, ACM, 2007.
- [17] S. Javanmardi, C. Lopes, and P. Baldi, "Modeling user reputation in wikis," *Statistical Analysis and Data Mining*, vol. 3, no. 2, pp. 126–139, 2010.
- [18] J. Leskovec, D. Huttenlocher, and J. Kleinberg, "Signed networks in social media," in *Proc. SIGCHI Conf. Human Factors in Computing Systems*, pp. 1361–1370, ACM, 2010.
- [19] F. B. Viégas, M. Wattenberg, and K. Dave, "Studying cooperation and conflict between authors with history flow visualizations," in *Proc. SIGCHI conf. Human factors in computing systems*, pp. 575–582, ACM, 2004.
- [20] A. Kittur, B. Suh, B. A. Pendleton, and E. H. Chi, "He says, she says: conflict and coordination in Wikipedia," in *Proc. SIGCHI conf. Human factors in computing systems*, pp. 453–462, ACM, 2007.
- [21] U. Brandes and J. Lerner, "Visual analysis of controversy in user-generated encyclopedias," *Information Visualization*, vol. 7, no. 1, pp. 34–48, 2008.
- [22] B. Suh, E. H. Chi, B. A. Pendleton, and A. Kittur, "Us vs. them: Understanding social dynamics in Wikipedia with revert graph visualizations," in *Proc. IEEE Symp. Visual Analytics Science and Technology (VAST'07)*, pp. 163–170, IEEE Computer Society, 2007.
- [23] R. Sumi, T. Yasseri, A. Rung, A. Kornai, and J. Kertész, "Edit wars in Wikipedia," in *Proc. 3rd intl. conf. Privacy, Security, Risk and Trust (PASSAT) and 3rd intl. conf. Social Computing (SocialCom)*, pp. 724–727, IEEE Computer Society, 2011.
- [24] T. Yasseri, R. Sumi, A. Rung, A. Kornai, and J. Kertész, "Dynamics of conflicts in Wikipedia," *PloS one*, vol. 7, no. 6, p. e38869, 2012.
- [25] M. Tsvetkova, R. García-Gavilanes, and T. Yasseri, "Dynamics of disagreement: Large-scale temporal network analysis reveals negative interactions in online collaboration," *Scientific reports*, vol. 6, p. 36333, 2016.
- [26] G. Wu, M. Harrigan, and P. Cunningham, "Characterizing Wikipedia pages using edit network motif profiles," in *Proc. 3rd intl. workshop Search and mining user-generated contents*, pp. 45–52, ACM, 2011.
- [27] J. Liu and S. Ram, "Who does what: Collaboration patterns in the Wikipedia and their impact on article quality," *ACM Transactions on Management Information Systems (TMIS)*, vol. 2, no. 2, p. 11, 2011.
- [28] J. Lerner and A. Lomi, "Knowledge categorization affects popularity and quality of Wikipedia articles," *PloS one*, vol. 13, no. 1, p. e0190674, 2018.
- [29] J. E. Blumenstock, "Size matters: word count as a measure of quality on Wikipedia," in *Proc. 17th Intl. Conf. WWW*, pp. 1095–1096, ACM, 2008.
- [30] S. Maniu, B. Cautis, and T. Abdesslem, "Building a signed network from interactions in Wikipedia," in *Proc. Databases and Social Networks*, pp. 19–24, ACM, 2011.
- [31] F. Flöck and M. Acosta, "WikiWho: Precise and efficient attribution of authorship of revised content," in *Proc. 23rd Intl. Conf. WWW*, pp. 843–854, ACM, 2014.
- [32] C. T. Butts, "A relational event framework for social action," *Sociological Methodology*, vol. 38, no. 1, pp. 155–200, 2008.
- [33] W. de Nooy, "Signs over time: Statistical and visual analysis of a longitudinal signed network," *Journal of Social Structure*, vol. 9, 2008.
- [34] U. Brandes, J. Lerner, and T. A. Snijders, "Networks evolving step by step: Statistical analysis of dyadic event data," in *Proc. 2009 Intl. Conf. Advances in Social Network Analysis and Mining (ASONAM)*, pp. 200–205, IEEE, 2009.
- [35] W. de Nooy, "Networks of action and events over time. A multilevel discrete-time event history model for longitudinal network data," *Social Networks*, vol. 33, no. 1, pp. 31–40, 2011.
- [36] W. de Nooy and J. Kleinnijenhuis, "Polarization in the media during an election campaign: A dynamic network model predicting support and attack among political actors," *Political Communication*, vol. 30, no. 1, pp. 117–138, 2013.
- [37] J. Lerner, M. Bussmann, T. A. Snijders, and U. Brandes, "Modeling frequency and type of interaction in event networks," *Corvinus Journal of Sociology and Social Policy*, vol. 4, no. 1, pp. 3–32, 2013.
- [38] M. Tsvetkova, R. García-Gavilanes, L. Floridi, and T. Yasseri, "Even good bots fight: The case of Wikipedia," *PloS one*, vol. 12, no. 2, p. e0171774, 2017.
- [39] Ø. Borgan, L. Goldstein, and B. Langholz, "Methods for the analysis of sampled cohort data in the Cox proportional hazards model," *The Annals of Statistics*, pp. 1749–1778, 1995.
- [40] D. Vu, P. Pattison, and G. Robins, "Relational event models for social learning in MOOCs," *Social Networks*, vol. 43, pp. 121–135, 2015.