Aalto University

OPEN ACCESS

This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.

Iniguez Gonzalez, Gerardo; Török, János; Yasseri, Taha; Kaski, Kimmo; Kertész, János

Modeling social dynamics in a collaborative environment

Published in:
EPJ Data Science

DOI:
10.1140/epjds/s13688-014-0007-z

Published: 01/01/2014

Document Version
Publisher's PDF, also known as Version of record

Published under the following license:
CC BY

Please cite the original version:
Iniguez Gonzalez, G., Török, J., Yasseri, T., Kaski, K., & Kertész, J. (2014). Modeling social dynamics in a collaborative environment. *EPJ Data Science*, *3*(7), 1-20. https://doi.org/10.1140/epjds/s13688-014-0007-z

This material is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.

Iñiguez et al. *EPJ Data Science* 2014, **2014**:7
http://www.epjdatascience.com/content/2014/1/7

EPJ Data Science
a SpringerOpen Journal

REGULAR ARTICLE Open Access

Modeling social dynamics in a collaborative environment

Gerardo Iñiguez[1], János Török[2], Taha Yasseri[3*], Kimmo Kaski[1] and János Kertész[2,4]

*Correspondence:
taha.yasseri@oii.ox.ac.uk
[3]Oxford Internet Institute, University of Oxford, Oxford, OX1 3JS, United Kingdom
Full list of author information is available at the end of the article

Abstract

Wikipedia is a prime example of today's value production in a collaborative environment. Using this example, we model the emergence, persistence and resolution of severe conflicts during collaboration by coupling opinion formation with article editing in a bounded confidence dynamics. The complex social behavior involved in editing articles is implemented as a minimal model with two basic elements; (i) individuals interact directly to share information and convince each other, and (ii) they edit a common medium to establish their own opinions. Opinions of the editors and that represented by the article are characterised by a scalar variable. When the pool of editors is fixed, three regimes can be distinguished: (a) a stable mainstream article opinion is continuously contested by editors with extremist views and there is slow convergence towards consensus, (b) the article oscillates between editors with extremist views, reaching consensus relatively fast at one of the extremes, and (c) the extremist editors are converted very fast to the mainstream opinion and the article has an erratic evolution. When editors are renewed with a certain rate, a dynamical transition occurs between different kinds of edit wars, which qualitatively reflect the dynamics of conflicts as observed in real Wikipedia data.

Keywords: social dynamics; mathematical modeling; peer-production; Wikipedia; bounded confidence; opinion dynamics; mass-collaboration; social conflict

1 Introduction

Cooperative societies are ubiquitous in nature [1], yet the cooperation or the mutual assistance between members of a society is also likely to generate conflicts [2]. Potential for conflicts is commonplace even in insect species [3] and so is conflict management through policing and negotiation in groups of primates [4, 5]. In human societies cooperation goes further not only in its scale and range, but also in the available mechanisms to promote conflict resolution [6, 7]. Collaborative and conflict-prone human endeavors are numerous, including public policy-making in globalized societies [8, 9], open-source software development [10], teamwork in operating rooms [11], and even long-term partnerships [12]. Moreover, information communication technology opens up entirely new ways of collaboration. With such a diversity in system size and social interactions between individuals, it seems appropriate to study this phenomenon of social dynamics in the framework of statistical physics [13, 14], an approach benefiting greatly from the availability of large scale data on social interactions [15, 16].

Springer

© 2014 Iñiguez et al.; licensee Springer. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited.

As a relevant example of conflicts in social cooperation we select Wikipedia (WP), an intriguing example of value production in an online collaborative environment [17]. WP is a free web-based encyclopedia project, where volunteering individuals collaboratively write and edit articles about any topic they choose. The availability of virtually all data concerning the visiting and editing of article pages provides a solid empirical basis for investigating topics such as online content popularity [16, 18] and the role of opinion-formation processes in a peer-production environment [19].

The editing process in WP is usually peaceful and constructive, but some controversial topics might give rise to extreme cases of disagreement over the contents of the articles, with the editors repeatedly overriding each other's contributions and making it harder to reach consensus. These 'edit wars' (as they are commonly called) result from complex online social dynamics, and recent studies [20] have shown how to detect and classify them, as well as how they are related to burstiness and what are their circadian patterns in editing activity [21].

Although collaborative behavior might appear without direct interactions between individuals, communication tends to have a positive effect on cooperation and trust [22]. Indeed, more immediate forms of communication (voice as opposed to text, for example) have been seen to increase the level of cooperation in online environments [23]. In WP, direct communication is implemented via 'talk pages', open forums where editors may discuss improvements over the contents of articles and exchange their related opinions [24]. Discussions among editors are not mandatory [25], but there is a significant correlation between talk page length and the likelihood of an edit war, indicating that many debates happen in articles and talk pages, simultaneously [17, 26].

Overall, a minimal model aimed at reproducing the temporal evolution of a common medium (i.e. a product collectively modified by a group of people, like an article in WP) requires at least the following two ingredients:

(i) *agent-agent dynamics*: Individuals share their views and opinions about changes in the article using an open channel accessible to all editors (the talk page or some other means of communication), thus effectively participating in an opinion-formation process through information sharing.

(ii) *agent-medium dynamics*: Individuals edit the article if it does not properly summarize their views on the subject, thus controlling the temporal evolution of the article and coupling it to the opinion-formation mechanism.

We describe the opinion-formation process taking place in the talk page by means of the well-known *bounded confidence* mechanism first introduced by Deffuant *et al.* [27], where real discussions take place only if the opinions of the people involved are sufficiently close to each other. Conversely, we model article editing by an 'inverse' bounded confidence process, where individuals change the current state of the article only if it differs too much from their own opinions. Particularly, we focus our attention on how the coupling between agent-agent and agent-medium interactions determines the nature of the temporal evolution of an article. This we consider as a further step towards the theoretical characterization of conflict in social cooperative systems such as WP [28].

The text is organized as follows: In Section 2 we introduce and discuss the model in detail. In Section 3 we describe our results separately for the cases of a fixed editor pool and a pool with editor renewal, and finally make a comparison with empirical observations on WP conflicts. In Section 4 we present concluding remarks.

2 Model

Let us first assume that there is a system of N agents as potential editors for a collective medium. The state of an individual i at time t is defined by its scalar, continuous opinion $x_i(t) \in [0,1]$, while the medium is characterized by a certain value $A(t)$ in that same interval. The variable x represents the view and/or inclination of an agent concerning the topic described by the common medium, while A is the particular position actually represented by the medium.

Although it may seem too reductive to describe people's perceptions by a scalar variable x, many topics can actually be projected to a one-dimensional struggle between two extreme, opposite options. In the Liancourt Rocks territorial dispute between South Korea and Japan [29], for example, the values $x = 0,1$ represent the extreme position of favoring sovereignty of the islets for a particular country. Other topics are of course multifaceted, generating discussions that depend on the global affinity of individuals and multiple cultural factors [30]. While this complexity could be tackled by the use of vectorial opinions [31, 32], our intention here is not to describe extremism as realistically as possible, but to study the rise of collaborative conflict even in the simplest case of binary extremism.

In the case of WP, the scalar variable A represents the opinion expressed by the written contents of an article, which carries the assumption that all agents perceive the medium in the same way. Real scenarios of public opinion might be more complex, given the tendency of individuals to attribute their own views to others and thus perceive false consensus [33], usually out of a social need to belong [34]. Even so, we consider A to be a sensible description of a WP article, one that could initially be measured by human judgment in the form of expert opinions, or in an automated way by quantifying lexical features and the use of certain language constructs. We note, however, that the actual value of A is not the main concern of our study. Instead, we are interested in how opinion differences in collaborative groups may eventually lead to conflict, specifically when such opinion differences are perceived with respect to a common medium that all individuals modify collectively.

2.1 Agent-agent dynamics

For the agent-agent dynamics (AAD) we consider a generic bounded-confidence model over a complete graph [27, 35], that is, a succession of random binary encounters among all agents in the system. We initialize every opinion $x_i(0)$ to a uniformly-distributed random value in the interval $[0,1]$. The initial medium value $A(0)$ is chosen uniformly at random from the same interval. This way, even an initially moderate medium $A \sim 1/2$ may find discord with extreme opinions at the boundaries. For each interaction we randomly select two agents i, j and compare their opinions. If the difference in opinions exceeds a given threshold ϵ_T nothing happens, but if $|x_i - x_j| < \epsilon_T$ we update as follows,

$$(x_i, x_j) \mapsto \big(x_i + \mu_T[x_j - x_i], x_j + \mu_T[x_i - x_j]\big). \tag{1}$$

The parameter $\epsilon_T \in [0,1]$ is usually referred to as the *confidence* or *tolerance* for pairwise interactions, while $\mu_T \in [0,1/2]$ is a *convergence* parameter. AAD is then a symmetric compromise between similarly-minded individuals: people with very different opinions simply do not pay attention to each other, but similar agents debate and converge their views by the relative amount μ_T.

The dynamics set by Eq. (1) has received a lot of attention in the past [13], starting from the mean-field description of two-body inelastic collisions in granular gases [36, 37]. Its final, steady state is comprised by $n_c \sim 1/(2\epsilon_T)$ isolated opinion groups that arise due to the instability of the initial opinion distribution near the boundaries. Furthermore, n_c increases as $\epsilon_T \to 0$ in a series of bifurcations [38]. In the limit $\mu_T \to 0$ corresponding to a 'stubborn' society, the asymptotically final value of n_c also depends on μ_T [39, 40]. The bounded-confidence mechanism has been extended in many ways over the years, considering interactions between more than two agents [41], vectorial opinions [31, 42–44], and coupling with a constant external field [45].

2.2 Agent-medium dynamics

For the agent-medium dynamics (AMD) we use what could be thought of as an asymmetric, inverse version of the bounded-confidence mechanism described above. When the opinion of a randomly chosen agent i is very different from the current state of the medium, namely if $|x_i - A| > \epsilon_A$, we make the update,

$$A \mapsto A + \mu_A[x_i - A], \tag{2}$$

where $\epsilon_A, \mu_A \in [0,1]$ are the tolerance and convergence parameters for AMD. In other words, individuals that come across a version of the medium portraying a radically different set of mind will modify it by the relative amount μ_A, where the threshold to define similarity is given by ϵ_A. Conversely, if $|x_i - A| < \epsilon_A$ we update,

$$x_i \mapsto x_i + \mu_A[A - x_i] \tag{3}$$

meaning that individuals edit the medium when it differs too much from their opinions, but adopt the medium's view when they already think similarly. Observe that the maximum meaningful value of μ_T is 1/2 (i.e. convergence to the average of opinions), while the maximum $\mu_A = 1$ implies changing the medium (opinion) so that it completely reflects the agent's (medium's) point of view.

The previous rules comprise our model for the dynamics of conflicts in WP given a *fixed agent pool*, that is, without agents entering or leaving the editing process of the common medium. In a numerical implementation of the model, every time step t consists of N updates of AAD given by Eq. (1) and of AMD following Eqs. (2) and (3), so that time is effectively measured in number of edits and the broad inter-event time distribution between successive edits (observed in empirical studies [20]) does not have to be considered directly. Given a fixed agent pool, AAD favors opinion homogenization in intervals of length $2\epsilon_T$ and can thus create several opinion groups for low tolerance, while AMD makes the medium value follow the majority group. Then, for a finite system there is a nonzero probability that any agent outside the majority group will be drawn by the medium to it, and the system will always reach consensus after a transient regime characterized by fluctuations in the medium value [28].

However, in real WP articles the pool of editors tends to change frequently. Some editors leave (due to boredom, lack of interest or fading media coverage on the subject, or are banned from editing by editors at a higher hierarchical level) and newly arrived agents do not necessarily share the opinions of their predecessors. Such feature of *agent renewal*

during the process or writing an article may destroy consensus and lead to a steady state of alternating conflict and consensus phases, which we take into account by introducing thermal noise in the model. Along with any update of AAD/AMD, one editor might leave the pool with probability p_{new} and be substituted by a new agent with opinion chosen uniformly at random from the interval $[0,1]$. The quantity $1/(Np_{\text{new}})$ then formally acts as the inverse temperature of the system, signaling a dynamical phase transition between different regimes of conflict [28].

3 Results

3.1 Fixed agent pool

In the presence of a fixed agent pool ($p_{\text{new}} = 0$) with finite size N, the dynamics always reaches a peaceful state where all agents' opinions lie within the tolerance of the medium. To show this, let us calculate the probability that an unsatisfied editor i changes the medium A for n consecutive times, such that afterwards $|x_i - A'| < \epsilon_A$ and the agent can finally stop its stream of edits. For fixed x_i and following Eq. (2), the final distance between editor and medium is $|x_i - A'| = (1-\mu_A)^n |x_i - A|$, so the inequality $|x_i - A'| < \epsilon_A$ is satisfied if $n > \ln \epsilon_A / \ln(1-\mu_A)$. The probability of agent i not participating in AAD for n time steps is $(1-2/N)^n$, while the probability of choosing it for AMD is $1/N^n$. Then the total probability of this stream of edits is $(1-2/N)^n/N^n$, which for large N and $\mu_A \sim 0$ might be very small, but always finite. After editor i gets into the tolerance interval of the medium, it will not perform additional edits and will eventually adopt the majority opinion close to the medium value. Similar events with other unsatisfied agents will finally result in full consensus and put an end to the dynamics.

The existence of a finite relaxation time τ to consensus (for finite systems) contrasts drastically with the behavior of the bounded confidence mechanism alone, where consensus is never attained for $\epsilon_T < 1/2$ [13]. In other words, the presence of agent-medium interactions promotes an agreement of opinions that would otherwise not exist in the agent-agent dynamics, even though it may happen after a very long time (i.e. $\tau \gg 0$). If we think of the medium as an additional agent with maximum tolerance (in the sense that it always interacts with the rest no matter what) and against which agents have a different tolerance ϵ_A (as opposed to ϵ_T), this result is reminiscent of previous observations for a bounded-confidence model with heterogeneous thresholds [35, 46]. There, even a small fraction of 'open-minded' agents with relatively high tolerance may bridge the opinion difference between the rest of the agents and lead to consensus.

In order to analyze all possible typical behaviors of the fixed agent pool dynamics, we perform extensive numerical simulations in systems of size ranging from $N = 10$ to 10^4, letting the dynamics evolve for a maximum time $\tau_{\max} = 10^4$. We then characterize the temporal evolution of medium and agent opinions as a function of ϵ_T, ϵ_A and μ_A, while keeping $p_{\text{new}} = 0$ for all results in this section. Finally, since the value of μ_T has no major effect other than regulating the convergence time of AAD [39, 40], from now on we fix it to the maximum value $\mu_T = 1/2$ in order to speed up the simulations as much as possible.

A sample time series of medium and agent opinions is shown in Figure 1. As a function of the medium convergence μ_A the temporal evolution of the system shows three distinctive behaviors. In regime **I** where μ_A is typically very small (Figure 1(A) and (D)), there is one or more 'mainstream' opinion groups near $x \sim 1/2$ with a majority of the agents in the system, and a number of smaller, 'extremist' opinion groups at positions closer to the boundaries

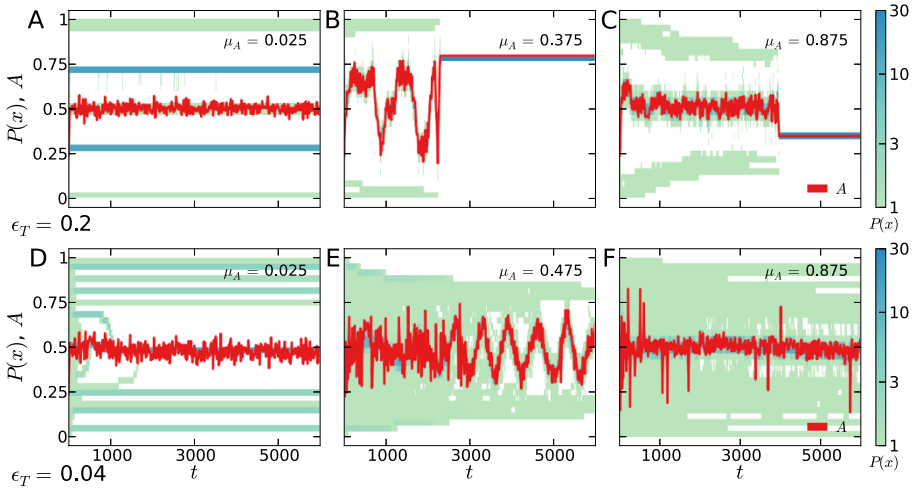

Figure 1 Temporal evolution of opinions and medium. (A, B, C) Time series of both the density distribution $P(x)$ of agents' opinions x (color map) and the medium value A (line) for $\epsilon_T = 0.2$ and several μ_A values, signaling the three different regimes found in the dynamics. **(D, E, F)** The same but for $\epsilon_T = 0.04$. Simulations correspond to $\epsilon_A = 0.1$ and $N = 10^4$.

$x = 0, 1$. The medium opinion stays on average at the center of the opinion space, close to the mainstream group(s), and although continuously contested by editors with extremist views, it remains stable and leads to a very slow convergence towards consensus. The reason for a long relaxation time in regime **I** is intuitively clear: for low μ_A any change in AMD is small and thus both medium and extremist opinions fail to converge quickly. When the tolerance ϵ_T decreases the effect is even more striking; even though the number of opinion groups is larger (according to the approximation $n_c \sim 1/[2\epsilon_T]$), the article is quite stable and remains close to the mainstream view.

In regime **II** identified with intermediate values of μ_A (Figure 1(B) and (E)), the fixed pool dynamics produces quasi-periodic oscillations in the medium value A, which appear after an initial stage of opinion group formation and end up very quickly in total consensus. Quite surprisingly, the final consensual opinion is not $x \sim 1/2$ (as in regime **I**) or that of the initial mainstream group, but some intermediate value closer to the extremist groups at the boundaries. This is indicative of a symmetry-breaking transition: as μ_A increases, a symmetric stationary state at $x \sim 1/2$ is replaced by a final state close to 0 or 1. The oscillations in regime **II** can initially be understood as a struggle over medium dominance among the different opinion groups created by AAD. The AMD mechanism couples the medium dynamics with these groups, exchanging agents between them and thus modifying their positions, until the majority group wins over the rest and consensus is achieved. For small ϵ_T oscillations are more well-defined and last for longer, while extremist groups tend to diffuse towards the mainstream.

In regime **III** for large μ_A (Figure 1(C) and (F)), extremist agents directly converge to a mainstream group and an article at the center. Since in this case μ_A is so large, after any jump of the article extremist agents can enter its tolerance interval and start drifting inwards. The limiting condition for this behavior is $\mu_A = 1 - \epsilon_A/(1/2 - \epsilon_A)$ [28], a line separating regime **III** from the rest. A smaller ϵ_T value produces a more erratic medium evolution, with occasional jumps up and down.

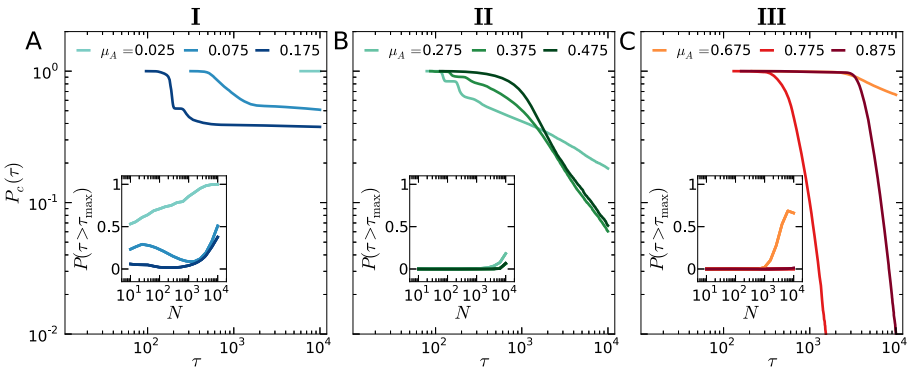

Figure 2 Distributions of relaxation time. (A, B, C) Cumulative distribution $P_c(\tau)$ of the relaxation time τ necessary to reach consensus and thus end the dynamics, for different values of the medium convergence μ_A. Insets: Probability $P(\tau > \tau_{\max})$ that the relaxation time is larger than $\tau_{\max} = 10^4$ (the maximum time allowed in the numerical simulations), as a function of N for selected values of μ_A. The symbols **I**, **II** and **III** denote the three different regimes found in the dynamics. Simulations correspond to $\epsilon_T = 0.2$, $\epsilon_A = 0.1$ and $N = 10^4$, with averages over 10^4 realizations.

The regimes of the fixed agent pool dynamics can be quantified on average by taking a look at the cumulative distribution $P_c(\tau)$ of the relaxation time τ (Figure 2). In regime **I** the tail of $P_c(\tau)$ is quite flat, getting flatter as μ_A decreases. In contrast, the distribution has a power-law and an exponential tail in regimes **II** and **III**, respectively, signaling shorter relaxation times. The only exception is the transition between **II** and **III**, where τ might be as large as in **I**. Since $P_c(\tau)$ tends to be broad, the average value of τ is not very meaningful and we opt instead for the probability $P(\tau > \tau_{\max})$ that the relaxation time is larger than a fixed maximum time. Numerically, we estimate $P(\tau > \tau_{\max})$ as the fraction of realizations of the dynamics that have not reached consensus after $\tau_{\max}$ time steps, out of a large total of 10^4 realizations. In regimes **II** and **III**, $P(\tau > \tau_{\max})$ remains small as N increases, indicating that τ is roughly independent of system size. On the other hand, $P(\tau > \tau_{\max})$ scales with N for **I** and for the boundary between **II** and **III**, even reaching 1 for appropriate values of μ_A and N. A corollary is that even modestly-sized systems may only reach consensus after an astronomical time, if the medium convergence value is appropriate.

The transition between regimes becomes even clearer when we consider the effect of the medium tolerance ϵ_A, resulting in a phase diagram for $P(\tau > \tau_{\max})$ in (ϵ_A, μ_A) space (Figure 3(A)). It turns out that regimes **I** and **II** cover most of the low ϵ_A values, while the line $\mu_A = 1 - \epsilon_A/(1/2 - \epsilon_A)$ roughly signals the transition to regime **III**, which covers a broad area of large ϵ_A. As N increases, the transition to **I** from either **II** or **III** (Figure 3(B) and (C)) becomes sharper: a consensual final state reached after a very short time gives way to a stationary state that remains stable for really long times. Such features of the phase diagram remain qualitatively unchanged if we substitute $P(\tau > \tau_{\max})$ with another measure giving robust statistics, such as the median relaxation time of the dynamics.

Finally, we can consider the symmetry-breaking transition between regimes **I** and **II** by taking a look at the density distribution $P(A)$ of the final medium value (Figure 4(A) and (B)). After either τ or $\tau_{\max}$ has passed, the majority of opinions are in consensus with A, making $P(A)$ a good approximation for the final opinion distribution $P(x)$ as well. In regime **I** the medium distribution is roughly unimodal and peaked at $A \sim 1/2$, signaling a stable and moderate medium. Here the relaxation time is quite long and for most realizations

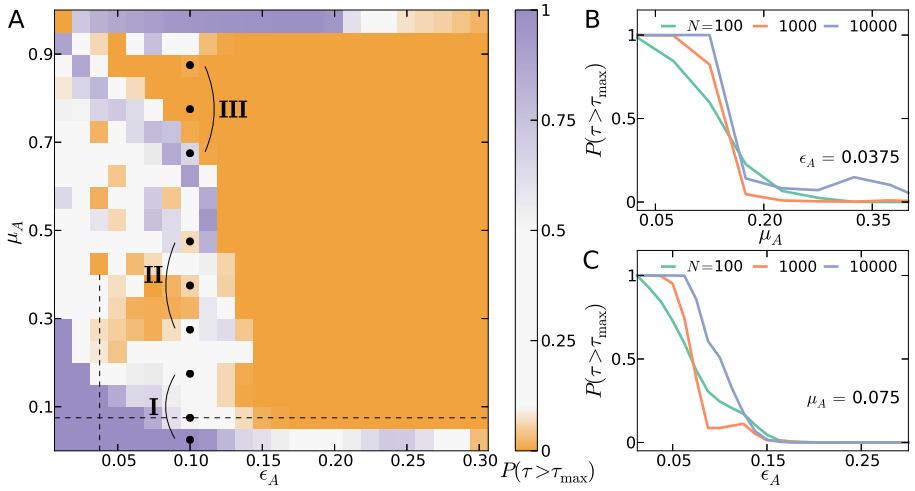

Figure 3 Phase diagram for a fixed agent pool. (A) Phase diagram in (ϵ_A, μ_A) space of the probability $P(\tau > \tau_{\max})$ that the relaxation time is larger than $\tau_{\max} = 10^4$, in a system of size $N = 10^4$. Points give the (ϵ_A, μ_A) values used in Figure 2, corresponding to regimes **I**, **II** and **III**. **(B, C)** Cross sections of the phase diagram along the dashed lines in **(A)** for varying N. Simulations correspond to $\epsilon_T = 0.2$, with averages over 10^4 realizations.

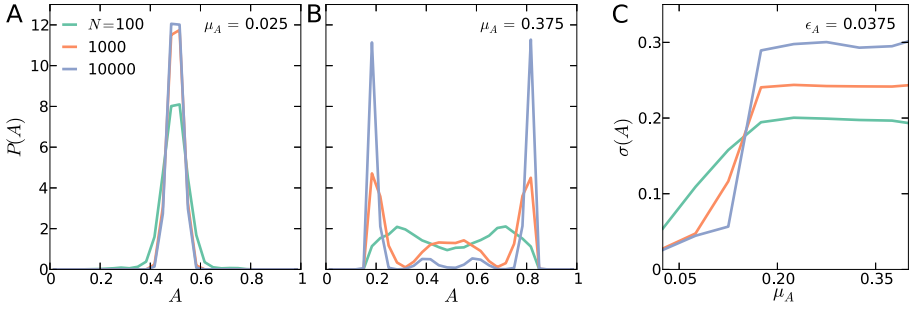

Figure 4 Symmetry-breaking transition. (A, B) Distribution $P(A)$ of the final medium value A reached after a time $\min(\tau, \tau_{\max})$ has elapsed, for varying N. The selected μ_A values represent regimes **I (A)** and **II (B)**. **(C)** Standard deviation $\sigma(A)$ of the final medium value as a function of μ_A, for several values of N. This order parameter signals a symmetry-breaking transition between regimes **I** and **II**. Simulations correspond to $\epsilon_T = 0.2$, $\epsilon_A = 0.0375$ and $\tau_{\max} = 10^4$, with averages over 10^4 realizations.

$\tau > \tau_{\max}$. In regime **II**, however, $P(A)$ becomes bimodal, meaning that the medium is more likely to end up close to the extremes rather than in the center. When N is large, the main peaks in $P(A)$ correspond to consensual final states with $\tau \leq \tau_{\max}$, while the secondary peaks are made up of long-lived realizations with long relaxation time. Larger values of $\tau_{\max}$, although computationally expensive, would therefore let us see a strictly bimodal medium distribution for regime **II**. As N increases the distribution peaks become sharper and we can use the standard deviation $\sigma(A)$ of the final medium value as an order parameter for the transition (Figure 4(C)). In the thermodynamic limit $N \to \infty$, a vanishing $\sigma(A)$ for **I** implies a stationary stable state with $A \sim 1/2$ and no consensus. As μ_A increases this symmetry gets broken, $\sigma(A)$ becomes nonzero and a true final state of consensus appears.

This symmetry-breaking mechanism may be understood analytically via a rate equation formalism [28]. The resulting rate equation can be solved numerically assuming three ed-

itor groups: a mainstream at $x \sim 1/2$ and two extremists with opinions close to the boundaries. The solution shows stability for the medium at the mainstream opinion when μ_A is small, but becomes unstable and oscillating for $\mu_A \simeq 3\epsilon_A \pm 0.1$. The bifurcation transition is very sensitive on the position of the extremists, depending not only on (μ_A, ϵ_A) but also on the initial conditions. This is in part the cause of the 'noisy' landscape of regime **II** in Figure 3(A), which appears regardless of the measure used to draw the phase diagram.

3.2 Agent renewal

In real systems the pool of collaborators is usually not fixed: Editors come and go and very often the number of editors fluctuates in time as external events may incite more or less attention. To keep things simple we only focus on systems with a fixed number of editors (N agents), but we allow agent replacement with probability $p_{\text{new}} \neq 0$. In our numerical simulations this happens prior to editing, and new agents have initially random opinions coming from a uniform distribution.

If $\epsilon_A < 1/2$ there is always an opinion range outside the article tolerance region $[A - \epsilon_A, A + \epsilon_A]$ and new agents may enter such range and edit the article. From WP data we know that even peaceful articles have few disputes now and then so such a scenario is realistic. This is thus in contrast with the case of a fixed opinion pool, where consensus is theoretically always achieved.

A stronger statement can be shown [28], namely that if

$$\epsilon_A > \epsilon_A^* = \frac{1}{2 - \mu_A} \tag{4}$$

then consensus is always reached after a finite number of steps, but if $\epsilon_A < \epsilon_A^*$ there are realizations that do not reach consensus ever. We show here an example: if the medium value is $A = \epsilon_A^*$, then for $\epsilon_A = \epsilon_A^* - \varepsilon$ an editor at $x = 0$ will disagree with the article and change it by $\Delta = \epsilon_A^* \mu_A$, so the new medium value would be $A = 1 - \epsilon_A^*$. Afterwards an agent at $x = 1$ can restore the article to its previous state and avoid consensus.

The lack of full consensus does not mean that the system is always in a conflict state. There are periods when A remains unchanged and these peaceful times are ended by conflicts in which the opinion of the article is continuously disputed between agent groups of different opinion. If the dispute is settled (i.e. all agents are satisfied by the article) a new peaceful period may start. The ratio of these peaceful and conflicting periods changes with the parameters and may be considered as a good candidate for an order parameter. Thus we define the order parameter P as the relative length of the peaceful periods.

The order parameter is plotted for two different initial conditions in Figure 5. The top figure shows the value of the order parameter P for a 'peaceful' initial condition when all agents had the opinion $x_i = 1/2$. The bottom figure was instead obtained for a system with 'conflict' initial conditions, namely one with 20% of agents divided between two extremist groups of opinions 0 and 1 (and the rest at $x_i = 1/2$) before the start of the dynamics.

It is clear that there are two distinct regimes in the phase diagram of Figure 5: one characterized by $P = 1$ ('peaceful' regime), the other with $P = 0$ ('conflict' regime) and a sharp transition in between. There is a region which is different in the two cases and will be discussed later. We then identify the transition point with the largest gradient of P by using the lower plot in Figure 5. The resulting phase diagram is shown in Figure 6.

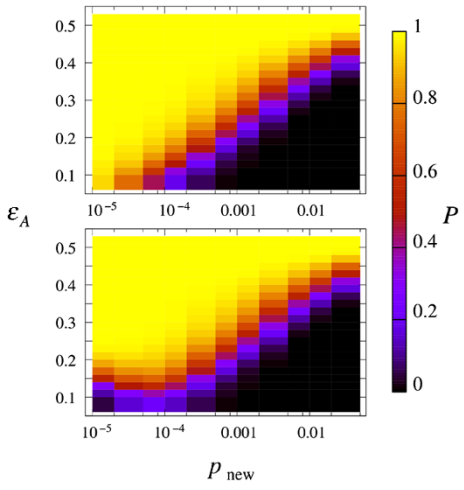

Figure 5 Order parameter for the agent renewal case. Order parameter P as a function of the medium tolerance ϵ_A and the agent replacement rate p_{new} for systems of size $N = 80$. The chosen initial condition is consensus for the top diagram and conflict for the bottom one.

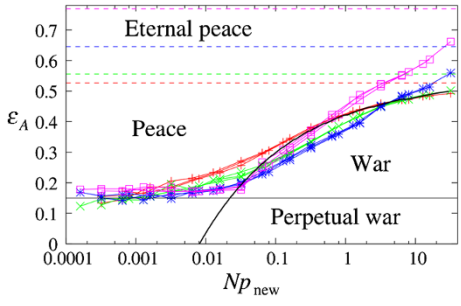

Figure 6 Phase diagram for a system with agent renewal. Largest gradient of P by using the lower plot in Figure 5, for varying μ_A. Simulations correspond to $\epsilon_T = 0.2$ and system size ranging from $N = 10$ to 640. The article convergence parameter was $\mu_A = 0.1, 0.2, 0.45, 0.7$ for red, green, blue and magenta respectively. The curved black line is the analytical result for $\mu_A = 0.1$. The horizontal line limiting the prepetual peace domain is at $\epsilon_A = 0.15$. The eternal peace is limited by ϵ_A^* (shown with dashed lines for the same color) which depends on μ_A.

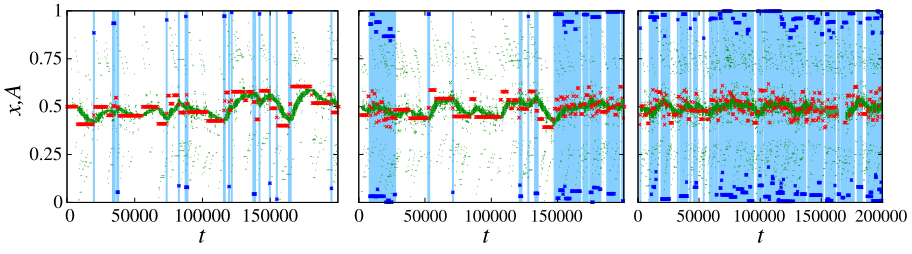

Figure 7 Time evolution of opinions. Samples of medium/agent opinions as a function of time for $\epsilon_A = 0.42$, and for three different regimes represented by $p_{\text{new}} = 0.001, 0.002, 0.003$ (from left to right respectively). Colour coding is as follows: Red points (opinion of the article), green dots (agents who are satisfied with the article), blue points (agents whose opinion is outside the medium tolerance interval), and light blue background (conflict regions).

This transition is further illustrated in Figure 7 where we display sample time evolutions of the opinions of agents and medium. The left panel shows an example of a peaceful regime. As mentioned before, from time to time new agents arrive with incompatible views with respect to the article but they are pacified very fast, i.e. the conflict periods are short. In the transition regime (middle panel) the scenario of peaceful times interrupted by short conflicts is still observable, but periods of continuous conflict occasionally appear. In the conflict regime exemplified by the right panel, these conflict bursts become persistent and the peaceful periods tend to disappear.

The above transition is the result of a competition between two timescales. New agents arrive outside of the article's tolerance interval with an 'insertion' timescale $\tau_{\text{ins}} \propto Np_{\text{new}}$. In order to have $P > 0$ the conflicts must be resolved before a new extremist agent arrives. Let us note that the convergence is very fast if there is only one extremist group. The problem is solved by displacing the article opinion by the required amount, which can be done in few (N independent) steps. This is what happens in the left panel of Figure 7. On the other hand, if we have two extremist groups on both sides of the opinion interval the relaxation is much slower and this is manifested in a much longer relaxation time. Thus, at the transition the insertion timescale is equal to the relaxation time of the case with two extremist groups, which is analogous to the fixed agent pool version of the model.

The task here is to determine the relaxation time of the fixed pool version of the model and relate it to τ_{ins}. For large values of the medium tolerance ($\epsilon_A > 1/4$), the relaxation time can by calculated analytically [28],

$$\tau(e) = c(\mu_A)N\big(\big[2e^2 + e_0^2(n-1)\big]n - ee_0(n-1)(2+n)\big), \tag{5}$$

where $e = \epsilon_A^* - \epsilon_A$, $e_0 = \epsilon_A^* - 1/2$, n denotes the integer part of e/e_0 (which is actually the number of steps the medium can make in one direction) and c is a constant depending on μ_A.

The above approach works well for $\epsilon_A > 0.3$ and $\mu_A < 0.3$ (regime **III** of the fixed pool case). If the mainstream group gets dissatisfied either by the large jump (μ_A is too large) or by the small tolerance (ϵ_A too small) of the article, the reasoning presented in [28] breaks down and new effect comes into play, namely the relaxation times of the fixed pool system becomes be enormous (regime **I**).

As we enter regime **I** of the fixed pool dynamics the relaxation time increases sharply (see Figure 3(B) and (C)). This means that if the system gets into a conflict state it will remain there for ever, which happens for,

$$\epsilon_{A,\text{lim}} = \frac{1}{4} - \frac{\epsilon_T}{2}. \tag{6}$$

This is why, starting from a conflict initial condition, the lower phase diagram in Figure 5 shows $P = 0$ for $\epsilon_A < 0.15$. On the other hand, in order to initiate such a conflict one needs to have a situation where two extremists appear on both ends of the opinion space outside of the article tolerance interval. If we have a single extremist then the consensus will be reached within a few time steps, independently of N. So the probability that we create a long-lasting conflict state decreases proportionally to the agent replacement probability. This is why we see only peace on the finite-time realizations leading to the upper phase diagram in Figure 5. Had we waited long enough, a conflict would have been formed for $\epsilon_A < 1/4 - \epsilon_T/2$ and would have persisted further on.

In summary, the typical behavior of our model in the presence of agent renewal may be divided into four distinct regimes:

(i) *Eternal peace* ($\epsilon_A > \epsilon_A^*$): The system reaches consensus very fast and remains there for ever.

(ii) *Peace* ($\epsilon_A > \frac{1}{4} - \frac{\epsilon_T}{2}$ and above the phase transition line): The system is mainly in a consensual state and only interrupted by short disputes.

(iii) *War* ($\epsilon_A > \frac{1}{4} - \frac{\epsilon_T}{2}$ and below the phase transition line): The system is mainly in a state of disagreement.

(iv) *Perpetual war* ($\epsilon_A < \frac{1}{4} - \frac{\epsilon_T}{2}$): In this regime and in the thermodynamic limit $N \to \infty$ no consensus may exist.

3.3 The case of Wikipedia

Although the model described and analyzed above is simplified enough to be extendable to various cases of collaboration, we specially intend to use it to explain some of the empirical observations regarding edit wars in WP.

Wikipedia is huge, not only in its number of articles and users but in the number of times articles are edited. In most cases articles are not written in a collaborative way, i.e., they have single authors or a few authors who have written and edited different parts of the article without any significant interaction [47]. In contrast, a few cases show significant constructive and/or destructive interactions between editors. The latter situation has been named 'edit war' by the WP community and defined as follows: "An edit war occurs when editors who disagree about the content of a page repeatedly override each other's contributions, rather than trying to resolve the disagreement by discussion" [48].

To start an empirical analysis of such opinion clashes and the way they are entangled with collaboration, we need to be able to locate and quantify edit wars.

3.3.1 Controversy measure

An algorithm to quantify edit wars and measure the amount of social clashes for WP articles has been introduced and validated before [49], and then used to study extensively the dynamical aspects of WP conflicts [20]. An independent study [50] has also shown that this measure is among the most reliable in capturing very controversial articles.

We quantify the 'controversiality' of an article based on its edit history by focusing on 'reverts' (i.e. when an editor completely undoes another editor's edit and brings the article back to the state just before the last version). Reverts are detected by comparing all pairs of revisions of an article throughout its history, namely by comparing the MD5 hash code [51] of the revisions. Specifically, a revert is detected when two versions in the history line are exactly the same. In this case the latest edit (leading to the second identical revision) is marked as a revert, and a pair of editors, referred to as reverting and reverted editors, are recognized.

Very soon in our investigation we noticed that reverts can have different reasons and not in all cases signalize a conflict of opinions. For example, an editor could revert personal edit mistakes or someone else's. Reverts are also heavily used to suppress vandalism, in itself a different type of destructive social behavior, but with no collaborative intention and therefore out of our interest. Thus we narrowed down our analysis to 'mutual reverts'. A mutual revert is recognized if a pair of editors (x, y) is observed once with x as the reverter and once with y. We also noticed that mutual reverts between pairs of editors at different levels of expertise and experience in WP editing could contribute differently to an edit war. Two experienced editors getting involved in a series of mutual reverts is usually a sign of a more serious conflict, as opposed to the case when two newbies or a senior editor and a newbie bite each other [52]. As a solution we introduced a 'weight' for each editor, and to sum up all reverts within the history of an article we counted each revert by using the smaller weight of the pair of editors involved in it. The weight of an editor x

is defined as the number of edits performed by him or her, and the weight of a mutually reverting pair is defined as the minimum of the weights of the two editors. The controversiality M of an article is then defined by summing the weights of all mutually reverting editor pairs, multiplying this number by the total number of editors E involved in the article. Overall,

$$M = E \sum_{\text{all mutual reverts}} \min\left(N^{\mathrm{d}}, N^{\mathrm{r}}\right), \tag{7}$$

where N^{r}, N^{d} are the number of edits for the article committed by the reverting/reverted editor. This measure can be easily calculated for each article, irrespective of the language, size, and length of its history.

Before starting our discussion about the empirical dynamics of conflict and its comparison with theoretical results, a remark on the most controversial articles in WP. We have calculated M for all articles in 13 different languages, from the start of each language WP up to March 2010. In Table 1 we show the list of the top-10 most controversial articles. A more complete and detailed analysis of the lists of the most controversial WP articles in different languages and differences and similarities between them can be found elsewhere [53].

3.3.2 Dynamics of conflict and war scenarios

Measuring M can not only lead us to rank the articles based on their cumulative controversy measure, but also enables us to follow edit wars in time as they emerge and get resolved, by investigating the evolution of M as time passes and the article develops. In the top row of Figure 8 we show the time evolution of M for three different sample articles.

Based on the way M evolves in time, we may categorize almost all controversial articles into three categories:

(i) *Single war to consensus*: In most cases controversial articles can be included in this category. A single edit war emerges and reaches consensus after a while, stabilizing quickly. If the topic of the article is not particularly dynamic, the reached consensus holds for a long period of time (top left in Figure 8).

(ii) *Multiple war-peace cycles*: In cases where the topic of the article is dynamic but the rate of new events (or production of new information) is not higher than the pace to reach consensus, multiple cycles of war and peace may appear (top center in Figure 8).

(iii) *Never-ending wars*: Finally, when the topic of the article is greatly contested in the real world and there is a constant stream of new events associated with the subject, the article tends not to reach a consensus and M increases monotonically and without interruption (top right in Figure 8).

The empirical war scenarios described previously are in qualitative agreement with the theoretical regimes of our model in the case of agent renewal, as seen from both the sample time series in Figure 7 and the regimes of war and peace in the phase diagrams of Figure 5 and Figure 6. Unfortunately, the theoretical order parameter P is quite difficult to measure in real systems as editor opinions are not known. What we know instead is the controversy measure M of Eq. (7). As mentioned before, M counts conflict events (i.e. mutual reverts)

Table 1 List of the most controversial articles in different language WPs according to *M*.

	English	German	French	Spanish	Portuguese
1	George W. Bush	Croatia	Ségolène Royal	Chile	São Paulo
2	Anarchism	Scientology	Unidentified flying object	Club América	Brazil
3	Muhammad	9/11 conspiracy theories	Jehovah's Witnesses	Opus Dei	Rede Record
4	List of WWE personnel	Fraternities	Jesus	Athletic Bilbao	José Serra
5	Global warming	Homeopathy	Sigmund Freud	Andrés Manuel López Obrador	Grêmio Foot-Ball Porto Alegrense
6	Circumcision	Adolf Hitler	September 11 attacks	Newell's Old Boys	Sport Club Corinthians Paulista
7	United States	Jesus	Muhammad al-Durrah incident	FC Barcelona	Cyndi Lauper
8	Jesus	Hugo Chávez	Islamophobia	Homeopathy	Dilma Rousseff
9	Race and intelligence	Minimum wage	God in Christianity	Augusto Pinochet	Luiz Inácio Lula da Silva
10	Christianity	Rudolf Steiner	Nuclear power debate	Alianza Lima	Guns N' Roses

	Czech	Hungarian	Romanian
1	Homosexuality	Gypsy Crime	FC Universitatea Craiova
2	Psychotronics	Atheism	Mircea Badea
3	Telepathy	Hungarian radical right	Disney Channel (Romania)
4	Communism	Viktor Orbán	Legionnaires' rebellion & Bucharest pogrom
5	Homophobia	Hungarian Guard Movement	Lugoj
6	Jesus	Ferenc Gyurcsány's speech in May 2006	Vladimir Tismăneanu
7	Moravia	The Mortimer case	Craiova
8	Sexual orientation change efforts	Hungarian Far-right	Romania
9	Ross Hedvíček	Jobbik	Traian Băsescu
10	Israel	Polgár Tamás	Romanian Orthodox Church

Table 1 (*Continued.*)

	Arabic	Persian	Hebrew	Japanese	Chinese
1	Ash'ari	Báb	Chabad	Koreans in Japan	Taiwan
2	Ali bin Talal al Jahani	Fatimah	Chabad messianism	Korea origin theory	List of upcoming TVB series
3	Muhammad	Mahmoud Ahmadinejad	2006 Lebanon War	Men's rights	TVB
4	Ali	People's Mujahedin of Iran	B'Tselem	Internet right-wing	China
5	Egypt	Criticism of the Quran	Benjamin Netanyahu	AKB48	Chiang Kai-shek
6	Syria	Tabriz	Jewish settlement in Hebron	Kamen Rider Series	Ma Ying-jeou
7	Sunni Islam	Ali Khamenei	Daphni Leef	One Piece	Chen Shui-bian
8	Wahhabi	Ruhollah Khomeini	Gaza War	Kim Yu-Na	Mao Zedong
9	Yasser Al-Habib	Massoud Rajavi	Beitar Jerusalem F.C.	Mizuho Fukushima	Second Sino-Japanese War
10	Arab people	Muhammad	Ariel Sharon	GoGo Sentai Boukenger	Tiananmen Square protests of 1989

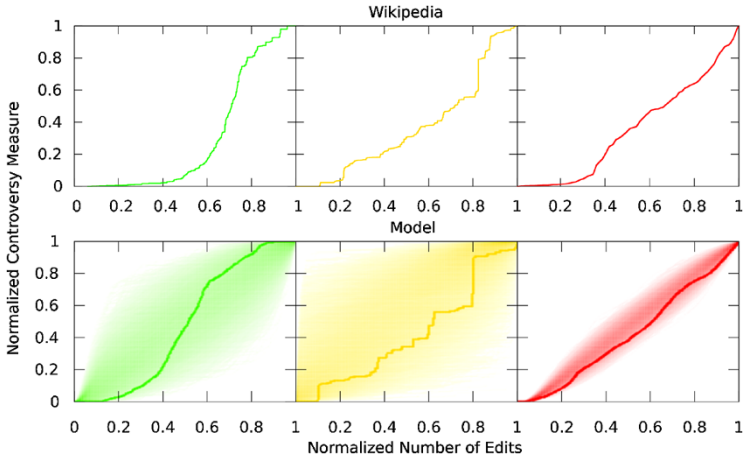

Figure 8 War scenarios for WP and our model. Top: Empirical controversy measure M as a function of the number of article edits in three different war scenarios. From left to right, sample articles are 'Jyllands-Posten Muhammad cartoons controversy', 'Iran', and 'Barack Obama'; and correspond to the regimes of 'single war', 'war-peace cycles', and 'never-ending war' respectively. Bottom: Theoretical conflict measure S in the case of agent renewal, reproducing the qualitatively analogue evolution of WP articles with parameter values $N = 640$, $\epsilon_T = 0.2$ and $\mu_A = 0.1$, as well as $\epsilon_A = 0.35, 0.42, 0.30$ and $p_{\text{new}} = 0.001, 0.001, 0.002$ for the three war scenarios, respectively. Continuous lines correspond to selected single runs of the model, while the shading indicates the density of S over an ensemble of 10^4 realizations.

and weights them by the maturity of the editor. This process can actually be repeated for the model: The editor maturity T_i is then defined as the number of time steps an agent has been in the pool of editors (a quantity constantly reset by agent replacement), and a conflict event is considered as the time an editor modifies the article, since this implies the agent is not satisfied with the state of the medium.

Thus a theoretical counterpart S to the WP controversy measure M may be defined as follows: Let $S = 0$ at the beginning of the dynamics. Then in each update t^* (out of the N that constitute a time step in the dynamics), when editor i changes the state of the article by the amount $\Delta = |A(t^* + 1) - A(t^*)|$ we increment S by $T_i\Delta$, where T_i measures the time i has been in the editorial pool. Examples of the temporal evolution of S (lower row in Figure 8) closely reproduce the qualitative behavior of M for different war scenarios. To further compare empirical observations in WP with our model predictions, we measure the typical length of a constant 'plateau' in the M and S time series, i.e. the number of edits between two successive increments. As seen in the distribution of plateau length for WP and the model (Figure 9), a statistical agreement for all three war scenarios is clear.

A last word on WP banning statistics. A way of estimating the number of extremists is to count the number of editors who have been 'banned' from editing. Explicitly, "a ban is a formal prohibition from editing some or all WP pages, either temporarily or indefinitely" [54]. Usually banning is used against vandals and/or editors who violate WP policies, especially those related to edit wars. In Table 2 we give some statistics of editors at different classes of editing activity, according to their number of edits. Interestingly, the relative population of banned editors is larger among more experienced editors (i.e. editors with more than 1000 edits). In other words, up to almost 20% of experienced editors could have been involved in edit wars. This is in complete accord with the choices we have made for

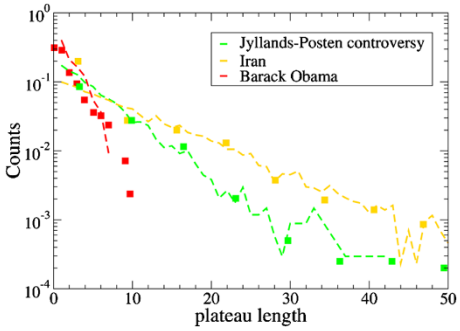

Figure 9 Peace periods in WP and our model. Distribution of plateau lengths for selected articles in WP (squares) and tuned parameters in our model (lines) for the three war scenarios shown in Figure 8. The length of a plateau or peace period is defined as the number of edits between two successive increments in either M or S.

the modeling setup, namely having two active extremist groups with roughly 20% of the total number of editors.

4 Discussion and conclusion

Here we have studied through modeling the emergence, persistence and resolution of conflicts in a collaborative environment of humans such as WP. The value production process takes place through interaction between peers (editors for WP) and through direct modification of the product or medium (an article). While in most cases this process is constructive and peaceful, from time to time severe conflicts emerge. We modeled the dynamics of conflicts during collaboration by coupling opinion formation with article editing in a generalized bounded-confidence dynamics. The simple addition of a common value-production process leads to the replacement of frozen opinion groups (typical of the bounded-confidence dynamics) with a global consensus and a tunable relaxation time. The model with a fixed pool shows a rich phase diagram with several characteristic behaviors: (a) an extremely long relaxation time, (b) fast relaxation preceded by oscillating behavior of the medium opinion, and (c) an even faster relaxation with an erratic medium. We have observed a symmetry-breaking, bifurcation transition between regimes (a) and (b), as well as divergence of the relaxation time in the transition between regimes (b) and (c).

If the pool is not fixed and editors are exchanged with new ones at a given rate, we obtain two different phases: conflict and peace. A conflict measure can be defined for the modeled system and be directly compared to its empirical counterpart in real WP data. It is then possible to follow the temporal evolution of this measure of controversy and obtain a good qualitative agreement with the empirical observations. These results lead us to plausible explanations for the spontaneous emergence of current WP policies, introduced to moderate or resolve conflicts.

Two remarks are at place here. In this study we have used a particular collaboration environment and compared our results with WP. The main reason behind is that for the free encyclopedia we have a full documentation of actions; however, we should emphasize that as web-based collaborative environments are abundant, we believe that our approach and results are much more general. Second, we are aware of the fact that the model contains a number of stringent simplifications: There are cultural differences between the WPs (e.g., in the usage of the talk page), and as in all human-related features there are large inhomogeneities in the opinions, in the tolerance level and in the activity of editors. Some of these aspects are under current study and will be taken into account for future research.

Table 2 Percentage of banned users to the total number of editors at three different classes.

	Num. editors w. >1,000 edits	Ban. editors w. >1,000 edits	% ban. w. >1,000 edits	Num. editors w. >100 edits	Ban. editors w. >100 edits	% ban. w. >100 edits	Num. editors w. >1 edits	Ban. editors w. >1 edits	% ban. w. >1 edits
English	36,280	6,114	0.17	189,174	20,342	0.11	4,552,685	403,851	0.09
German	8,714	1,561	0.18	34,777	3,458	0.10	511,291	31,996	0.06
French	5,286	694	0.13	21,940	1,700	0.08	394,385	16,681	0.04
Spanish	3,834	765	0.20	19,135	2,404	0.13	479,305	21,850	0.05
Portuguese	1,733	345	0.20	8,077	1,015	0.13	194,584	7,486	0.04
Czech	700	112	0.16	2,439	236	0.10	48,030	2,663	0.06
Hungarian	844	138	0.16	3,107	276	0.09	49,024	1,201	0.02
Romanian	437	53	0.12	1,675	130	0.08	36,631	914	0.02
Arabic	610	96	0.16	2,736	198	0.07	80,498	1,085	0.01
Persian	580	151	0.26	2,531	406	0.16	56,805	2,544	0.04
Hebrew	1,009	233	0.23	3,898	515	0.13	53,318	4,341	0.08
Japanese	4,010	786	0.20	19,090	2,845	0.15	242,621	21,995	0.09
Chinese	2,106	378	0.18	9,002	1,072	0.12	160,579	9,387	0.06

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
All authors designed the research and participated in the writing of the manuscript. GI and JT contributed equally to this work. GI and JT performed the numerical calculations and analytical approximations. TY analyzed the empirical data.

Author details
[1]Department of Biomedical Engineering and Computational Science, Aalto University School of Science, Aalto, FI-00076, Finland. [2]Institute of Physics, Budapest University of Technology and Economics, Budapest, H-1111, Hungary. [3]Oxford Internet Institute, University of Oxford, Oxford, OX1 3JS, United Kingdom. [4]Center for Network Science, Central European University, Budapest, H-1051, Hungary.

Acknowledgements
The authors acknowledge support from the ICTeCollective EU FP7 project. JK thanks FiDiPro (TEKES) and the DATASIM EU FP7 project for support. JT thanks the support of European Union and the European Social Fund through project FuturICT.hu (grant no.: TAMOP-4.2.2.C-11/1/KONV-2012-0013).

Received: 17 March 2014 Accepted: 18 July 2014 Published online: 24 September 2014

References

1. Axelrod R, Hamilton WD (1981) The evolution of cooperation. Science 211(4489):1390
2. Schelling TC (1980) The strategy of conflict. Harvard University Press, Cambridge
3. Ratnieks FLW, Foster KR, Wenseleers T (2006) Conflict resolution in insect societies. Annu Rev Entomol 51(1):581-608
4. de Waal FBM (2000) Primates—a natural heritage of conflict resolution. Science 289(5479):586-590
5. Flack JC, Girvan M, de Waal FBM, Krakauer DC (2006) Policing stabilizes construction of social niches in primates. Nature 439(7075):426-429
6. Melis AP, Semmann D (2010) How is human cooperation different? Philos Trans R Soc B 365(1553):2663-2674
7. Rand DG, Arbesman S, Christakis NA (2011) Dynamic social networks promote cooperation in experiments with humans. Proc Natl Acad Sci USA 108(48):19193-19198
8. Quirk PJ (1989) The cooperative resolution of policy conflict. Am Polit Sci Rev 83(3):905-921
9. Buchan NR, Grimalda G, Wilson R, Brewer M, Fatas E, Foddy M (2009) Globalization and human cooperation. Proc Natl Acad Sci USA 106(11):4138
10. Lerner J, Tirole J (2002) Some simple economics of open source. J Ind Econ 50(2):197-234
11. Rogers D, Lingard L, Boehler ML, Espin S, Klingensmith M, Mellinger JD, Schindler N (2011) Teaching operating room conflict management to surgeons: clarifying the optimal approach. Med Educ 45(9):939-945
12. Minson JA, Liberman V, Ross L (2011) Two to tango: effects of collaboration and disagreement on dyadic judgment. Pers Soc Psychol Bull 37(10):1325-1338
13. Castellano C, Fortunato S, Loreto V (2009) Statistical physics of social dynamics. Rev Mod Phys 81(2):591-646
14. Helbing D (2010) Quantitative sociodynamics: stochastic methods and models of social interaction processes, 2nd edn. Springer, Berlin
15. Onnela J-P, Saramäki J, Hyvönen J, Szabó G, Lazer D, Kaski K, Kertész J, Barabási A-L (2007) Structure and tie strengths in mobile communication networks. Proc Natl Acad Sci USA 104(18):7332-7336
16. Ratkiewicz J, Fortunato S, Flammini A, Menczer F, Vespignani A (2010) Characterizing and modeling the dynamics of online popularity. Phys Rev Lett 105(15):158701
17. Yasseri T, Kertész J (2013) Value production in a collaborative environment. J Stat Phys 151(3-4):414-439
18. Mestyán M, Yasseri T, Kertész J (2013) Early prediction of movie box office success based on Wikipedia activity big data. PLoS ONE 8(8):e71226
19. Ciampaglia G (2011) A bounded confidence approach to understanding user participation in peer production systems. In: Datta A et al (eds) Social informatics. Lecture notes in computer science, vol 6984. Springer, Berlin, pp 269-282
20. Yasseri T, Sumi R, Rung A, Kornai A, Kertész J (2012) Dynamics of conflicts in Wikipedia. PLoS ONE 7(6):e38869
21. Yasseri T, Sumi R, Kertész J (2012) Circadian patterns of Wikipedia editorial activity: a demographic analysis. PLoS ONE 7(1):e30091
22. Kollock P (1998) Social dilemmas: the anatomy of cooperation. Annu Rev Sociol 24(1):183-214
23. Jensen C, Farnham SD, Drucker SM, Kollock P (2000) The effect of communication modality on cooperation in online environments. In: Proceedings of the SIGCHI conference on human factors in computing systems. CHI'00. ACM, New York, pp 470-477
24. Wikipedia: Talk page guidelines. Retrieved Feb 23, 2014, from http://en.wikipedia.org/wiki/Wikipedia:Talk_page_guidelines
25. Wikipedia: Using talk pages. Retrieved Feb 23, 2014, from http://en.wikipedia.org/wiki/Wikipedia:Using_talk_pages
26. Kaltenbrunner A, Laniado D (2012) There is no deadline: time evolution of Wikipedia discussions. In: Proceedings of the eighth annual international symposium on Wikis and open collaboration. WikiSym'12. ACM, New York
27. Deffuant G, Neau D, Amblard F, Weisbuch G (2000) Mixing beliefs among interacting agents. Adv Complex Syst 3(4):87-98
28. Török J, Iñiguez G, Yasseri T, San Miguel M, Kaski K, Kertész J (2013) Opinions, conflicts, and consensus: modeling social dynamics in a collaborative environment. Phys Rev Lett 110(8):088701
29. Wikipedia: Liancourt Rocks dispute. Retrieved May 21, 2014, from http://en.wikipedia.org/wiki/Liancourt_Rocks_dispute
30. Axelrod R (1997) The dissemination of culture. A model with local convergence and global polarization. J Confl Resolut 41(2):203-226
31. Lorenz J (2007) Continuous opinion dynamics under bounded confidence: a survey. Int J Mod Phys C 18(12):1819-1838

32. Sznajd-Weron K, Sznajd J (2005) Who is left, who is right? Physica A 351(2):593-604
33. Wojcieszak M, Price V (2009) What underlies the false consensus effect? How personal opinion and disagreement affect perception of public opinion. Int J Public Opin Res 21(1):25-46
34. Morrison KR, Matthes J (2011) Socially motivated projection: need to belong increases perceived opinion consensus on important issues. Eur J Soc Psychol 41(6):707-719
35. Weisbuch G, Deffuant G, Amblard F, Nadal J-P (2002) Meet, discuss, and segregate! Complexity 7(3):55-63
36. Ben-Naim E, Krapivsky PL (2000) Multiscaling in inelastic collisions. Phys Rev E 61(1):R5-R8
37. Baldassarri A, Marini Bettolo Marconi U, Puglisi A (2002) Influence of correlations on the velocity statistics of scalar granular gases. Europhys Lett 58:14
38. Ben-Naim E, Krapivsky PL, Redner S (2003) Bifurcations and patterns in compromise processes. Physica D 183(3-4):190-204
39. Laguna MF, Abramson G, Zanette DH (2004) Minorities in a model for opinion formation. Complexity 9(4):31-36
40. Porfiri M, Bollt EM, Stilwell DJ (2007) Decline of minorities in stubborn societies. Eur Phys J B 57(4):481-486
41. Hegselmann R, Krause U (2002) Opinion dynamics and bounded confidence models, analysis, and simulation. J Artif Soc Soc Simul 5(3):2
42. Fortunato S, Latora V, Pluchino A, Rapisarda A (2005) Vector opinion dynamics in a bounded confidence consensus model. Int J Mod Phys C 16(10):1535-1551
43. Jacobmeier D (2005) Multidimensional consensus model on a Barabási-Albert network. Int J Mod Phys C 16(4):633-646
44. Lorenz J (2008) Fostering consensus in multidimensional continuous opinion dynamics under bounded confidence. In: Managing complexity: insights, concepts, applications. Springer, Berlin, pp 321-334
45. González-Avella JC, Cosenza MG, Eguíluz VM, San Miguel M (2010) Spontaneous ordering against an external field in non-equilibrium systems. New J Phys 12:013010
46. Lorenz J (2010) Heterogeneous bounds of confidence: meet, discuss and find consensus! Complexity 15(4):43-52
47. Kimmons R (2011) Understanding collaboration in Wikipedia. First Monday 16:12
48. Wikipedia: Edit warring. Retrieved Feb 23, 2014, from http://en.wikipedia.org/wiki/Wikipedia:Edit_warring
49. Sumi R, Yasseri T, Rung A, Kornai A, Kertész J (2011) Edit wars in Wikipedia. In: 2011 IEEE third international conference on privacy, security, risk and trust (PASSAT) and 2011 IEEE third international conference on social computing (SocialCom), pp 724-727
50. Sepehri Rad H, Makazhanov A, Rafiei D, Barbosa D (2012) Leveraging editor collaboration patterns in Wikipedia. In: Proceedings of the 23rd ACM conference on hypertext and social media. HT'12. ACM, New York, pp 13-22
51. Rivest RL (1992) The MD5 message-digest algorithm. Internet Request for Comments, RFC 1321
52. Halfaker A, Kittur A, Riedl J (2011) Don't bite the newbies: how reverts affect the quantity and quality of Wikipedia work. In: Proceedings of the 7th international symposium on Wikis and open collaboration. WikiSym'11. ACM, New York, pp 163-172
53. Yasseri T, Spoerri A, Graham M, Kertész J (2014) The most controversial topics in Wikipedia: a multilingual and geographical analysis. In: Fichman P, Hara N (eds) Global Wikipedia: international and cross-cultural issues in online collaboration. Scarecrow Press, Lanham
54. Wikipedia: Banning policy. Retrieved Feb 23, 2014, from http://en.wikipedia.org/wiki/Wikipedia:Banning_policy

doi:10.1140/epjds/s13688-014-0007-z
Cite this article as: Iñiguez et al.: **Modeling social dynamics in a collaborative environment.** *EPJ Data Science* 2014 **2014**:7.

Submit your manuscript to a SpringerOpen journal and benefit from:

► Convenient online submission
► Rigorous peer review
► Immediate publication on acceptance
► Open access: articles freely available online
► High visibility within the field
► Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com