

July 2009

Persistence of Cultural Norms in Online Communities: The Curious Case of WikiLove

Myshkin Ingawale

Author, myshkinonline@gmail.com

Rahul Roy

Indian Institute of Management, Calcutta, rahul.roy@iimcal.ac.in

Priya Seetharaman

Indian Institute of Management, Calcutta, priya.seetharaman@iimcal.ac.in

Follow this and additional works at: <http://aisel.aisnet.org/pacis2009>

Recommended Citation

Ingawale, Myshkin; Roy, Rahul; and Seetharaman, Priya, "Persistence of Cultural Norms in Online Communities: The Curious Case of WikiLove" (2009). *PACIS 2009 Proceedings*. 108.

<http://aisel.aisnet.org/pacis2009/108>

This material is brought to you by the Pacific Asia Conference on Information Systems (PACIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in PACIS 2009 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

PERSISTENCE OF CULTURAL NORMS IN ONLINE COMMUNITIES: THE CURIOUS CASE OF WIKILOVE

Myshkin Ingawale

Doctoral Student
Indian Institute of Management, Calcutta
DH Road, Joka, Kolkata, WB 700104, India
myshkinonline@gmail.com

Rahul Roy

Professor of Management Information Systems
Indian Institute of Management, Calcutta
DH Road, Joka, Kolkata, WB 700104, India
rahul@iimcal.ac.in

Priya Seetharaman

Assistant Professor of Management Information Systems
Indian Institute of Management, Calcutta
DH Road, Joka, Kolkata, WB 700104, India
priyas@iimcal.ac.in

Abstract

Tremendous progress in information and communication technologies in the last two decades has enabled the phenomenon of Internet-based groups and collectives, generally referred to as online communities. Many online communities have developed distinct cultures of their own, with accompanying norms. A particular research puzzle is the persistence and stability of such norms in online communities, even in the face of often exponential growth rates in uninitiated new users. We propose a network-theoretic approach to explain this persistence. Our approach consists of modelling the online community as a network of interactions, and representing cultural norms as transmissible ideas (or 'memes') propagating through this network. We argue that persistence of a norm over time depends, amongst other things, on the structure of the network through which it propagates. Using previous results from Network Science and Epidemiology, we show that certain structures are better than others to ensure persistence: namely, structures which have scale-free degree distributions and assortative mixing. We illustrate this theory using the case of the community of contributors at Wikipedia, a collaboratively generated online encyclopaedia.

Keywords: Online Communities, Cultural Norms, Wikipedia, Social Production

INTRODUCTION

In the last two decades, the growth of the Internet and communication technologies has facilitated the rapid emergence of online interactions of dispersed groups of people with shared interests. This phenomenon of Internet-based groups and collectives, generally referred to as online or virtual communities is characterized by individuals using new technologies, such as the Internet, to fulfil not only economic but also social goals. In a seminal work, Rheingold (1993) describes online communities as “social aggregations that emerge from the Internet when enough people carry on public discussions long enough, with sufficient human feeling, to form webs of personal relationships” and suggests that they are replacing public spaces such as pubs and cafes as loci of public social interaction. As in any “face-to-face” community in the physical world, there is a history of events, constellation of values, a common lingo – a set of cultural norms – that many online communities develop. The focus of this work is to develop a satisfactory explanation for how such cultural norms survive and *continue to persist* in the online community, despite often extreme turbulence in the form of new uninitiated users joining at near-exponential rates.

EXPLAINING THE EMERGENCE OF CULTURAL NORMS

Norms are the agreed-upon expectations and rules by which a culture guides the behaviour of its members in any given situation (CliffsNotes.com, 2009). The question of why norms develop in the first place has interested researchers for a long time, since well before the new phenomenon of online communities, and literature in the areas of sociology and anthropology offers many interesting and often differing perspectives. In much of sociology literature, the concepts of “community” and “culture” share a very close relationship: A community is defined by its characteristic culture and it is cultural norms that are particularly influential in shaping the life of community members. Life Course Theory (Giele and Elder, 1998) suggests that the existence of norms is advantageous for the community, since it provides a guide to decision-making in an otherwise very complex environment. Epstein (2001) points out an extreme case: *Conformity to a norm without even thinking about it*. Psychological advantages of normative reasoning may also explain why explicit sanctions are not needed to enforce norms in all situations. For instance, Heckhausen (1999), states that cultural norms may have been internalized in our society, rendering obsolete any need for external societal enforcement. Another explanation for the emergence of cultural norms is based on Common Ground Theory, which provides a framework for understanding how two people or a small group develop *shared understanding* in a conversation (Clark & Brennan, 1991). A different, game-theoretic approach rooted in agent-based computational economics suggests the necessity of emergence of norms for the evolution of cooperation in society (Axelrod 1986). A well-established area of research, with roots in classical social psychology, is concerned with how physical proximity, similarity of beliefs and attitudes, amount of interaction, and affective ties are interrelated. For example, Carley (1991) has developed network models of how interacting individuals influence each other to produce homogeneity of beliefs. An excellent review of the culture-cognition-networks intersection is provided by Kilduff and Corley (2000). A related perspective is that of embeddedness (Granovetter, 1985), which suggests that the long run development of cultural norms is the result of collective dynamics within a social network. This perspective portrays individuals as being in the situation of taking a decision at the micro level under guidance of norms imposed at the macro level. In turn, the interactive dynamics of all micro level decisions within a certain society or community generates the macro level behaviour of the system which may either strengthen the existing norms or weaken them if there is a collective trend to deviate.

RESEARCH GAP : THE PUZZLE OF PERSISTENCE

While several theories have emerged to explain cultural norms in communities in the real ‘physical’ world, can they be ported over to cultural norms in online communities? We need to be aware that, in comparison to traditional communities, online communities differ in several respects. Traditional communities are place-based and members are rooted in geography. Group dynamics often override individual expression. There is a distinct border between membership and differentiation, that is, it is clearly defined who is a member and who is not: For example one cannot “subscribe” to become a

member of the Tutsi community of Rwanda in the same way, or with as much ease as one can to become a member of an online community based around eBay, Slashdot, RedCafe or Wikipedia. In contrast, online communities exist according to identification to an idea or task, rather than place. They are organized around an activity, and they are formed as a need arises (Squire & Johnson, 2000). All this makes an important point: Communities in the physical world usually have entry barriers in place for newcomers and are therefore *less* subject to change and turbulence due to the influx of new community members than online communities. Consequently, *persistence* is not a major issue for theories that have emerged in the literature to explain cultural norms. However, we see that for online communities, where exponential growth in number of users is often not un-usual, this is definitely a major issue. Cultural norms in communities require social interaction over a period of time to be internalized. Constant change and turbulence in community membership would entail a perpetual deviation from the status-quo, without stability of any kind. We would have not culture, but *chaos*. And yet, we observe a distinct and relatively stable culture and set of norms at many online communities. Even as the membership of such communities is undergoing rapid change, the distinct community culture often persists. Existing literature does not provide a clear answer: Why do cultural norms persist, why is *culture* stable, in many online communities characterized by turbulent change?

WIKIPEDIA AND THE CURIOUS CASE OF WIKILOVE

An example of the above mentioned puzzle of persistence of norms is to be found also in the Wikipedia community. Wikipedia is a popular web-based encyclopaedia edited freely and collaboratively by its users. Users may register but registration is not compulsory to edit articles. Since its inception, Wikipedia has grown near-exponentially in key metrics such as number of editors and number of articles. Articles are written by a process of concurrent contributions by volunteer unpaid editors, of different ages, nationalities, ethnic and professional backgrounds, with nothing specific in common except internet access and a motivation to contribute to the encyclopaedia. Within the Wiki publication model, are embedded the processes of conflict, negotiation and collaboration. The MediaWiki platform on which Wikipedia runs enables these processes but *consensus building* is by no means guaranteed: The shape and form an article takes is entirely up to the contributors and the interactions that occur between them. These interactions take the form of user responses and comments to each change made to the article (which are logged in the history) and discussions on separate dedicated pages. Besides these, a more ‘subtle’ interaction is implicit and inherent in the Wiki publication model itself: Consider the basic use-case - “User edits article”. In doing so, the user is implicitly establishing an interaction tie with a) The previous contributor of the article whose contribution was modified and b) All the other previous contributors of the article, whose contributions, by virtue of not being modified, received a tacit acceptance and approval. Figures 1 and 2 show sample interactions that occur in response to changes made on the Wikipedia article “Political Status of Taiwan”. As is apparent from the article’s sensitive topic, consensus is by no means guaranteed. To use a contextual metaphor - The interacting parties may not be *on the same page*, even after several rounds of discussions!

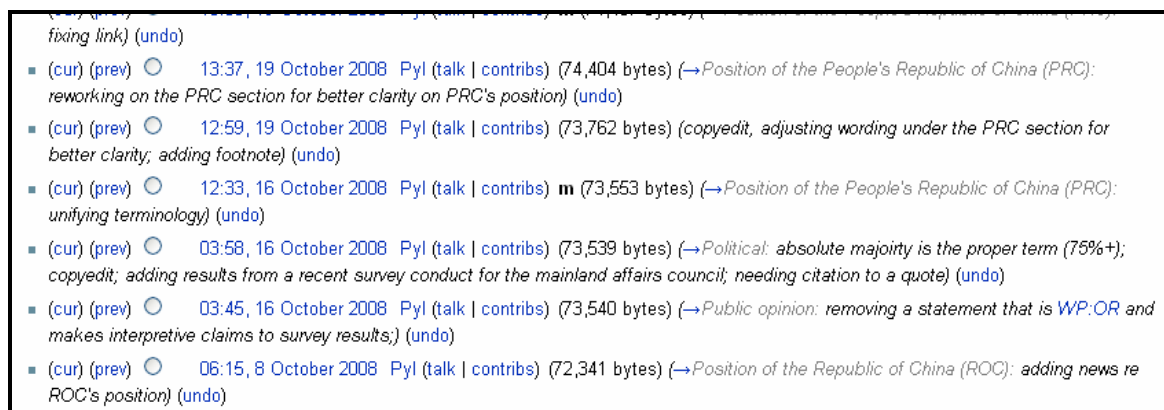


Figure 1. Revision History sample for Article on “Political Status of Taiwan”

<p>KMT is called Chinese Nationalist Party [edit]</p> <p>"When the Kuomintang visited Mainland China in 2005, the government-controlled Chinese media called this event as a "visit," and called the KMT one of "Taiwan's political parties" even though the Kuomintang's full name remains the "Chinese Nationalist Party."</p> <p>This is not true: CCTV, a government-controlled media, broadcast this visit and called KMT 'Zhongguo Guomindang', even they emphasize the 'Zhongguo' pronunciation http://www.tudou.com/programs/view/pZ7TNOBDWlw/</p> <p>Hand15 (talk) 07:47, 18 January 2008 (UTC)</p>
<p>China's own contradiction [edit]</p> <p>I personally made several changes in that section. I believe most of the content in this section is due to the lack of understanding. C&C welcomed.--StrikeEagle 02:40, 1 February 2007 (UTC)</p> <p>The entire article was originally written by pro-independence supporters , who seems not interested in clearly and concisely presenting the Pro-China POV. Redcloud822 18:17, 11 April 2007 (UTC)</p>
<p>Slip of the tongue [edit]</p> <p>Is this section necessary? All those cases were just as the title indicates "slip of the tongue"; compared to official stands those incidents don't carry any weight. It seems a bit childish to use "slip of the tongue" as arguments to support a particular POV. I suggest this section to be removed. I also added the fact that PRC treats Hongkong investment as "foreign investment" as well. Redcloud822 06:27, 25 March 2007 (UTC)</p>

Figure 2. Discussion Page sample for Article on “Political Status of Taiwan”

Despite the apparent impossibility of the task of achieving consensus, on an average, consensus *is* achieved, a common ground *is* reached, and the article is all the better for it. There is evidence for this in the form of numerous academic studies that have praised Wikipedia for its surprisingly high quality, neutral and credible *encyclopaedic* content. (Chesney, 2006). It has been pointed out that one of the main contributing factors is “WikiLove” (Reagle, 2007) - a general spirit of collegiality and mutual understanding, that has persisted in Wikipedia since the time of its inception in 2001. The WikiLove principle encapsulates both culture and governance, and is both implicit and explicit in what exactly it constitutes. The WikiLove page entry on Wikipedia lists some key components: “Follow Wikiquette - respect other contributors”, “Love newcomers even more”, “Follow Wikipedia policies—they make it easier to work with one another”, “Assume good faith, and assume the assumption of good faith”, “Aim for a neutral point of view—write articles that people from all sides can read and agree with”, “Stay cool—don't react hastily in anger. Instead, take some distance if you're feeling mad”, etc. A discussion extract such as the one below illustrates WikiLove in action:

"Cheers on your good work on Clothing! Since you seem knowledgeable on the subject, could you explain what you meant by (...) I read it, but the meaning wasn't clear from the context. Perhaps you could make it more understandable to the layman, or provide some links for further study? Thanks!"

WikiLove constitutes – or at least very strongly influences – most values, maxims and guidelines for behaviour and performing different tasks, and binds all *Wikipedians* – editors of Wikipedia – together into a community. This community has been described as one of the most vibrant communities on the internet (Kolbitsch and Maurer, 2006). That this community has a specific culture and shared norms of its own is further evidenced by the specialized language it has developed (with words like “to wikify”, “giffie”, “sock puppet”, “wikiholiday”, etc) that is meaningful only within the Wikipedia community context. Repeated interaction between users leads to the transmission of these norms across Wikipedia, from user to user, and from article to article. For instance, the norm “Aim for a neutral point of view” gets transmitted to a new user by repeated interaction and dialogue with more experienced editors, as well as through policy pages. The norm in action via the agency of a user results in an article written with a neutral point of view. New users read such articles, and take pointers. As they edit other articles, the “neutral point of view” norm gets passed on to these articles as well. In this way, two sets of entities – articles and users – interact amongst themselves in different ways and transmit norms across the interaction network. The persistence of such norms over time, is one of the major factors for Wikipedia’s robust governance mechanism and remarkably high quality of content – for its *stability* and continued existence as an encyclopaedia. This persistence is highly surprising when one considers that Wikipedia’s contributors and articles are growing near-exponentially. There is an influx of new users (see Figure 3) who are uninitiated in the ways of Wikipedia. They have yet to internalize its culture - They are not born with WikiLove. And yet, WikiLove lives on, as evidenced by the longevity of cultural norms such as “Aim for a neutral point of view” .

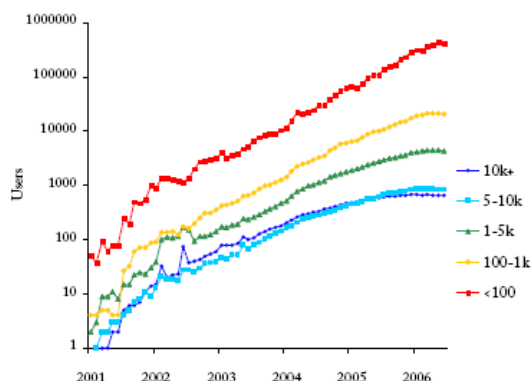


Figure 3. Near-exponential user count growth (shown here on logarithmic scale) of different categories of Wikipedia users, ordered by number of edits. Source: Kittur et al(2007)

This is all the more surprising in the light of Kolbitsch and Maurer’s (2006) study of online communities and the ‘Dunbar’s number’ which suggests that if a group is larger than approximately 150 individuals, it will become less individualistic and will require more restrictive regulations to provide governance. This has not happened to Wikipedia: In fact, it is difficult to imagine a successful forced imposition of restrictive regulations on Wikipedia, for the simple reason that contributors are not paid employees. At any time, they can just stop contributing. What binds them together as Wikipedians are not economic ties, but ties of *WikiLove*. Therefore, how *WikiLove lives on* even as a new generation of contributors takes over Wikipedia, how cultural norms *persist* in online communities, in spite of the community membership itself undergoing turbulent change, is an interesting research question.

SIMULATING WIKILOVE CONTAGION

To better understand and visualize the situation as it plays out in Wikipedia and other online communities, we use a simulation based approach. To this end, we need a) A model of the network of interactions within the community and b) A useful representation of a cultural norm.

The methodological approach is based on the premises of *embeddedness* and theories of *contagion* (Molina, 1995). The premise of embeddedness, much discussed since Granovetter’s (1985) pioneering work, is interpreted here in the context of the network of interactions within the community of contributors in Wikipedia. We use the dataset of the Cebuano¹ Wikipedia to build the network. Each community member is a node, and ‘connected’ to all other members he/she has interacted with. In the case of Wikipedia, we use the notion of ‘subtle’ interaction as discussed above. Users are represented as nodes and two users are ‘connected’ if they have contributed to a common article. We refer to this as the ‘User Network’. Similarly, in the corresponding ‘Article Network’, every article is a node, and two articles are ‘connected’ if they share at least one common contributor. Thus we get 2 representations of the network of interactions. To represent ‘cultural norm’ we use the concept of ‘meme’, first introduced by Dawkins (1976), who coined the term to describe a ‘socially transmissible idea’. He proposed that the meme is to culture what the gene is to biology, a replicator – “a unit of information with the ability to reproduce itself using resources from some material substrate”– that serves as a basis for the transmission of culture. Any idea that can be so transmitted can be a meme, but only those memes which are universally accepted acquire the status of “cultural norm”. The spread of memes is modelled as a function of inter-nodal transmission along interaction ties. Ties are

¹ Cebuano is a language spoken in the Philippines. The use of this comparatively smaller Wikipedia over say, the English Wikipedia has the dual advantages of being easier to process, and easier to visualize in one screenshot and printed page. Cebuano Wikipedia meta-revision-history XML dataset is available at <http://download.wikimedia.org/> (accessed 29th September 2008). It contains data in the form “contributor_name” made edit to “article_name” at “timestamp”, which we use to visualize the interaction network in NetLogo 4.2 development environment.

conceived of as conduits or paths along which information or influence flow. Seen from the point of view of the community as a whole, nodes are mutually influencing and informing each other via repeated interactions. The ultimate distribution of memes, which our simulation is interested in finding, is a function of the structure of the underlying interaction network. In this context, *persistence* of a particular meme can be equated to how *resilient* it is to competing memes. For examples the meme “Aim for a neutral point of view” competes with the meme “Present only your point of view, even if biased” to become the dominant meme i.e. a cultural norm. In this simulation, when we talk of a node to have been ‘infected’ by a meme, we mean the node has adopted that meme. Thus, in our case, if we use the language of epidemiology, a meme which infects a significant proportion of the network, and persists stably as an ‘epidemic’ for a long time, is a cultural norm. The simulation is made to run for 100 time steps i.e. *turns*. Infectiousness (I) is the probability of a node being infected with a particular meme at each time step. Chance-of-Recovery (R) is the probability of an infected node from recovering from the infection at each time step. For visual intuition, we colour code ‘infected’ nodes as yellow crosses, to differentiate them from the un-infected nodes, represented by dots (see Figure 5).

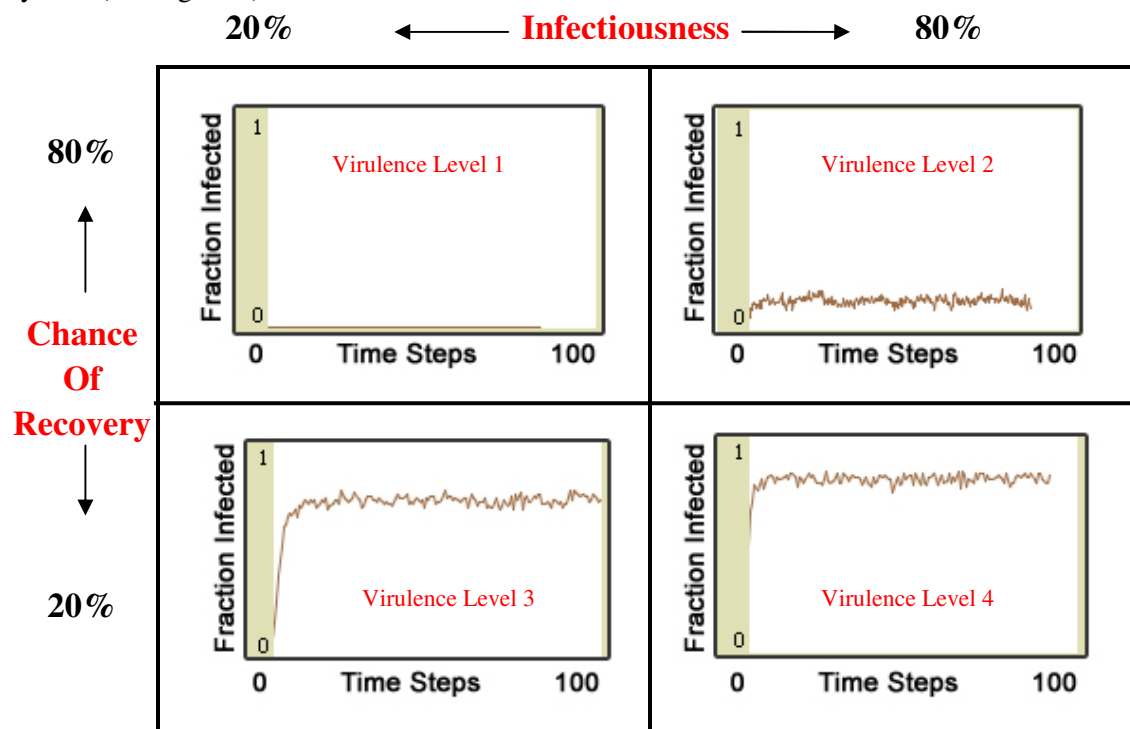


Figure 4. Simulation: Representative ‘Successful’ Runs of Responses of the Wikipedia Article Network to memes with different values of Infectiousness and Chance-of-Recovery

At every turn, an infected node has a probability I of passing on the meme to any of its neighbours and a probability R of recovering from the infection. With reference to our example, this means an article with at least one contributor who has adopted the “Aim for a neutral point of view” meme has a probability I of making one of its neighbouring articles in the interaction network adopt the neutrality meme, and a probability R of ‘recovering’ i.e. forsaking the neutrality meme (of course, via the agency of an editing user). For the sake of exhaustiveness, we consider four cases: High I , High R ; High I , Low R ; Low I , High R ; and Low I , Low R . (Low = 20% and High = 80%). This corresponds to memes with 4 discrete “levels of virulence”. Obviously, extreme cases are of a meme with Virulence Level 4: High I , Low R being a viral ‘idea’ that catches on quickly but is hard to let go of; to the other extreme, a meme of Virulence Level 1: Low I , High R , being an idea that is not very infectious, and is also easy to let go of even when once infected. For each of these 4 levels of virulence, we perform 25 runs each i.e we infect one article at random at each turn for 25 turns with the same setting of I and R . The first thing to notice, of course, is that the network’s probability of getting infected varies across the 4 cases. This makes intuitive sense, as a meme with high

infectiousness and giving its hosts less chance of recovery would be more likely to lead to an ‘epidemic’ i.e infect a majority of the nodes. Not all runs in which a meme is introduced lead to an epidemic. However, when the infection *does* start, we observe that it persists stably for the full 100 time steps interval of the experiment, in a dynamic equilibrium. The response of the system is non-linear: Either the infection spreads and attains a steady state or it doesn’t and completely dies out. Figure 4 displays the representative steady state “success” scenarios for the meme “Aim for a neutral point of view” (i.e scenarios where the infection has taken off), showing the fraction of nodes in the largest connected component² that have been infected. We notice that except for the case of meme of Virulence Level 1, which does not infect the network in any of the 25 runs, all other cases yield a steady state, a constant fraction of nodes infected, in a particular fraction of the runs (Table 2).

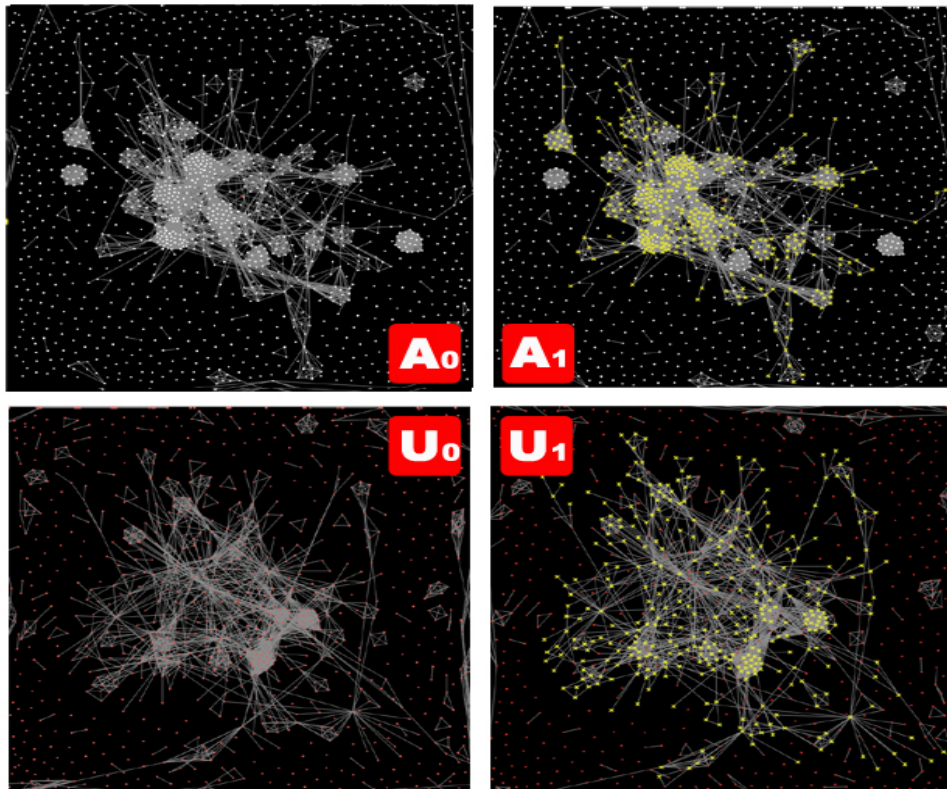


Figure 5. A_0, U_0 : Actual Cebuano Wikipedia Article and User Network Visualizations;
 A_1, U_1 : Simulation of Cebuano Wikipedia Article and User Networks with “Aim to be Neutral” meme (infected nodes are represented by Yellow ‘x’ s)

Consider the steady state where the “Aim to be neutral” meme is persisting over time within this network. We now introduce a competitor – ‘challenger’ meme – “Present only your point of view, even if biased” and allow it to propagate through the network. The results are very interesting. We find that the success rate of this new meme is very low. We had set this meme’s virulence level at 4, the maximum, i.e it has a 80% infectiousness and a 20% chance of recovery – and yet it has great difficulty in ‘taking over’ the network from the incumbent memes, of Virulence Levels greater than 1, as evidenced by very low success rates (except for the last case in Table 2, where the incumbent meme was absent in the network). Figure 6 shows a representative run in which it *does* manage to take over the network from an incumbent meme of Virulence Level 3. Importantly, the new meme, highly virulent as it may be, has *no effect* in cases where the incumbent meme itself is very virulent.

² A component is a connected chain of nodes. The largest connected component is often referred to as the “Giant Component”. The formation of the Giant Component is called *Percolation*

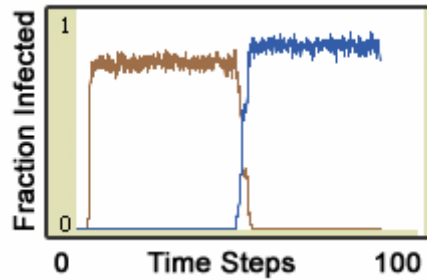


Figure 6. Rare Instance of Article Network with Incumbent Meme (brown line) of Virulence Level 3 being Taken over by New Meme (blue line) of Virulence Level 4

The system thus shows signs of ‘path-dependence’ i.e its final state is highly dependent on the initial conditions. The first meme, once incumbent in the network, is extremely difficult or sometimes impossible to dislodge, even by a highly virulent newcomer meme. In the context of the often quoted ‘Social Movement’ analogy of Wikipedia, it is not difficult to imagine the original ‘incumbent’ set of memes in Wikipedia to be highly virulent themselves– i.e both very infectious and difficult to let go of once infected. Hence, very few or no newly introduced memes stand a chance of dislodging the original ‘WikiLove’ memes. This observed *persistence* and stability of the network with the old memes, even in the face of the influx of new virulent competing memes, indicates that a part of the explanation for persistence of norms does indeed lie in the underlying *network structure*.

Virulence Level	Infectiousness (I)	Chance-of-Recovery (R)	Number of Runs	Number of Attempts Successful at Infecting Core	Probability of Epidemic that persists	Success Rate of New ‘Challenger’ Meme with Virulence Level 4
4	80%	20%	25	25	100%	0%
3	20%	20%	25	9	36%	4%
2	80%	80%	25	4	16%	8%
1	20%	80%	25	0	0%	100%

Table 2. Descriptive Statistics of Simulation Runs

A NETWORK-THEORETIC EXPLANATION

The simulation results indicate that the network *structure* is playing some role in the observed persistence. Basing our approach on pioneering work in the mathematics of assortative mixing in networks by Newman (2002), we focus on: **Degree Distribution** and **Degree Correlation**.

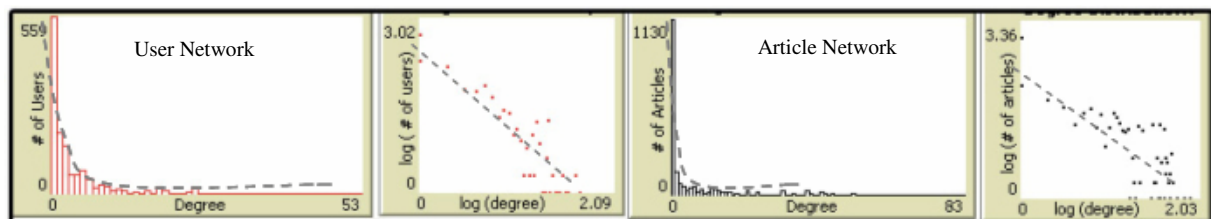


Figure 7. Evidence of Scale Free degree distribution in Wikipedia’s User and Article Networks.

Power Law [$P(k) \sim k^{-\gamma}$] : Exponent Estimation					
Cebuano Wikipedia’s User Network			Cebuano Wikipedia’s Article Network		
γ_u	R2	C.I *	γ_a	R2	C.I *
1.64	0.85	0.14	1.25	0.78	0.16

Table 3. Power Law Exponent γ in Cebuano Wikipedia’s User and Article Networks (* $\alpha = .12$)

1.1 Degree Distribution

The degree of a node is the count of its neighbours. The spread in node degree is characterized by a distribution function $P(k)$, which gives the probability that a randomly selected node has exactly k edges. Unlike a Normal Distribution, the Scale Free distribution observed in many real networks is characterized by the presence of a small number of nodes with a large number of links and a large number of nodes with a small number of links. It is formed by a ‘rich get richer’ mechanism whereby as new nodes are added, they attach themselves to old nodes with a probability proportional to the old node’s degree (Barabasi and Albert, 1999). We find evidence of **power laws** [$P(k) \sim k^{-\gamma}$] and **scale-free degree distribution** in both the Article and User Networks of Cebuano Wikipedia (See Figure 7 and Table 3 above). This implies that there are a few ‘hubs’ – high degree nodes – with a disproportionately large number of edges, while the majority of nodes only have a few edges.

1.2 Degree Correlation

Degree Correlation is the propensity of nodes of similar degree to connect to each other preferentially (Newman, 2002). Common examples of networks with degree correlation are science co-authorship networks (a scientist with many publications would be more likely to collaborate with another high-publication scientist than with any randomly picked scientist). Degree correlation is defined by the following equation

$$k_{nn}(k) = \sum_{k'} k' \Pr(k' | k)$$

$\Pr(k' | k)$ is the conditional probability that the vertex with degree k is adjacent to the vertex with degree k' . According to this property, networks are said to exhibit assortative mixing (or positive correlation) if nodes of a given degree tend to be attached with higher likelihood to nodes with similar degree. The network is said to exhibit disassortative mixing (or negative correlation) if the nodes of a given degree tend to be attached with higher likelihood to nodes with dissimilar degrees i.e. high degree nodes attaching preferentially to low degree nodes. We found strong positive degree correlation i.e. **assortative mixing** in Wikipedia’s Article Network. The log-log plot of degree k against the average degree of neighbouring nodes (Figure 8) is a straight line that fits the equation $\log(k_{nn}) = 0.58 \log(k) + 0.57$, with $R^2 = 0.85$. This approximates to $k_{nn} \sim k^{0.58}$. Thus, we find the degree correlation exponent as $\beta = 0.58$, and the Pearson’s Correlation Coefficient $r = 0.92$, which is an extremely high positive relationship. To put this in context, r for co-authorship networks in biology and physics has been found to be 0.127 and 0.363 respectively.

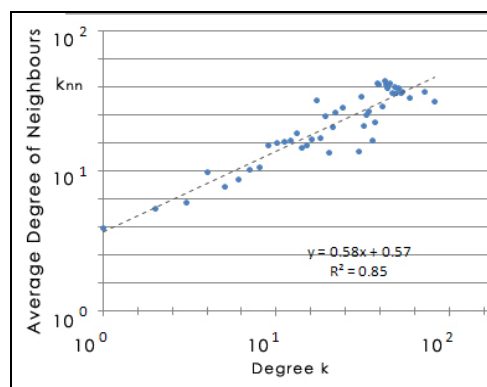


Figure 8. Assortative Mixing in Cebuano Wikipedia’s Article Network

Newman (2002) found that assortative networks percolate³ more easily and that they are also more robust to removal of their highest degree vertices. These two aspects of scale-free assortative networks lend themselves to the cultural norms’ persistence in the following way:

³ Percolation is the formation of a giant component – the largest connected chain of nodes.

1.1 Resilience to Failure by Random Removal and Targeted Attack

Scale-Free networks are extremely resilient to random removal of nodes. However, the connectivity of networks formed by disassortative mixing can be destroyed by the removal of just a few of the highest degree vertices. In *assortatively mixed networks*, however, removing high degree vertices is also a relatively inefficient strategy for destroying network connectivity, because these vertices tend to be clustered together in the core group, so there is redundancy in connectivity links. This is an important property because it implies that once attained, connectivity is seldom lost quickly in an assortatively mixed network, even if a few ‘key’ high-degree nodes are later removed. Interaction pathways are robust and resilient to both random failure as well as targeted removal of nodes, in assortatively mixed networks. Since the cultural norm sustains as a result of repeated interactions across nodes, the resilience of the interaction network is a necessary condition for persistence of norms

1.2 The Formation of a Reservoir

If the network mixes assortatively, then the high-degree vertices will tend to stick together in a sub-network or core group of higher mean degree than the network as a whole. Percolation occurs earlier within such a sub-network. The core group of an assortatively mixed network thus forms a “*reservoir*” for the virus/meme, sustaining an epidemic/culture even in cases in which the network is not sufficiently dense on average for the disease to persist. In an exponentially growing online community, a large number of new users, with very few interaction ties are added every time period. In spite of this, if the “reservoir” has been formed, the original cultural norms persist and eventually are adopted by the newcomers, as a result of repeated interactions with the core group of high degree nodes.

Consistent with the Simulation results, Wikipedia’s Interaction Network’s Structural Properties satisfy the conditions that would be required for the persistence of WikiLove over time. The results suggest that a Network-Theoretic approach may go some way towards providing a satisfactory explanation for the puzzle of persistence of cultural norms in many online communities.

LIMITATIONS

We realize there are certain limitations in the approach used. Firstly, the recourse to the ambiguous concept of *meme*, especially since the research community has been unable to arrive at a consensus on the definition and scope of the term. Secondly, the methodology used depends on the fundamental concept of capturing the *interaction network* as faithfully as possible – however, ‘interaction’ is a broad term and it is difficult to capture or even judge what constitutes *meaningful* interaction in the context of many online communities. Furthermore, an ‘interaction tie’ is not forever– in reality, it has a finite lifetime. This needs to be incorporated in the model. A related limitation is the lack of attention in this work to the *strength* of ties. Some ties might be more important than others – or the *type* of ties – positive or negative- may also matter. Lastly, *culture* is a complex and multifaceted phenomenon that cannot be explained away by our proposed network theoretic approach alone – There is a need to widen the scope and put this work in the larger context of discourses on culture in anthropology and sociology.

DISCUSSION AND SUMMARY

Reagle (2007) remarked that “While Wikipedia may very well work in practice, it can never work in theory”. This work has been an attempt at reconciling what has been observed to work in practice with existing theory. We proposed a network theoretic explanation for the puzzling phenomenon of *persistence* of cultural norms. We illustrated our points using the case of the Wikipedia community, and concluded that there are two structural requirements for the interaction network for persistence to occur : a) A Scale Free degree distribution and b) Assortative Mixing.

The theory behind this assertion, rooted in Newman’s (2002) work, emphasizes the assortative mixing process by which high-degree nodes link preferentially to other high-degree nodes. This, in

conjunction with the power law degree distribution, results in very early percolation and the formation of a dense, tightly connected inner core of nodes which acts as a ‘reservoir’ to sustain any early memes. While the interesting aspect is that the system is path-dependent, and the early meme is almost impossible to dislodge once it has infected the inner core, of course, the converse is that since percolation is restricted to this sub-network, the giant component of the network as a whole has a smaller size (as Figure 9 shows). Whereas in similar network studies of academic co-authorship networks in fields like biology and sociology, as well as teams in creative enterprises like Broadway Musicals (Guimera et al, 2005), a ‘tipping point’ had been observed beyond which the giant component emerged larger, engulfing almost all the network participants in its midst (“the invisible college” linking everyone in the domain), we notice that this is not the case in a strongly assortative network like Wikipedia. Perhaps, paradoxically, Wikipedia is more ‘elitist’ than Academia or Broadway : The co-authorship network and broad-way team network provided intra-disciplinary connectivity more or less uniformly to all network participants. In Wikipedia’s Article Network, however, presence of strong assortative mixing means that high degree nodes preferentially connect almost exclusively to other high degree nodes, forming a dense but small ‘giant component’ – which is by no means inclusive. The majority of nodes lie *outside* the giant component, in small fragmented clusters (See Figure 5 and 9). While the ‘closed club’ nature of the inner core of the giant component is responsible for the positive externality of providing stability and the much discussed *persistence* of cultural norms, and consequently, a reliable governance mechanism, it also raises an important question: *How long is this sustainable?*

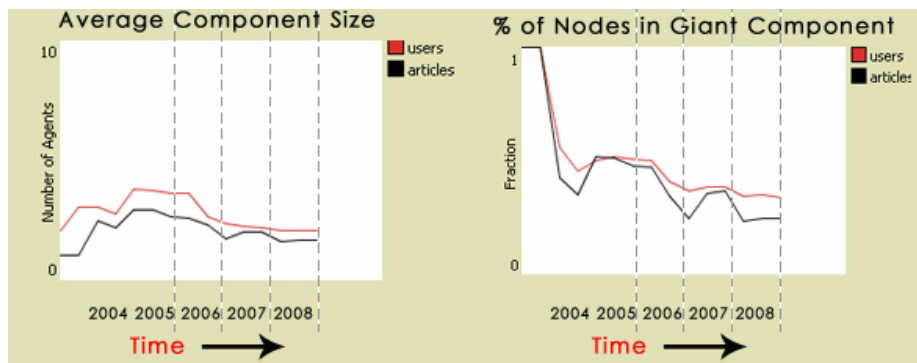


Figure 9. Average Component Size and Percentage of Nodes in the Giant Component

We see from Figure 9 that the percentage of nodes in the giant component is dropping over time i.e the proportion of newcomers – relative ‘outsiders’ to the closed club of the inner core that controls cultural norms is rising. With it, come its own set of problems. Kittur (2007) found that the ‘costs of collaboration’ are rising markedly in Wikipedia – i.e the cost of maintenance, spending time and effort discussing, undoing vandalism, etc. The interaction ties by which the inner core of nodes controls and coordinates Wikipedia are gradually getting strained. An interesting project at this stage in Wikipedia’s life-cycle might be to forecast the *breaking-point* of social production. Time necessarily spent arguing, discussing and doing administrative duties, rather than directly working on the content of the article itself, may well drive away some of the key committed ‘hub’ users – the custodians of WikiLove - who hold Wikipedia together at present.

This work may be of interest for businesses engaged in knowledge work, for whom Wikipedia remains an enticing puzzle. While management literature suggests that Knowledge Workers are autonomous and almost impossible to control (Drucker, 1999), here we have the world’s largest knowledge project, built almost entirely by the free labour of knowledge workers distributed all over the world. Governance by persistent cultural norms is an important part of the picture. In a broader context, understanding the conditions under which such persistence occurs should interest customer-centric businesses world-wide. There is a growing realization that online community members, bound together by a certain common culture, have a definite ROI for business. For instance, businesses are realizing that community users visit the corporate website nine times more often (McKinsey), remain customers 50% longer than (AT&T) and spend 54% more (EBay) than non-community users. Online

communities, bound by a common culture, also help users help themselves in a more satisfactory manner, while also reducing customer support costs for businesses (*statistics compiled*, Cothrel and Johnston, 2007). Our results suggest that these businesses should focus their energies on identifying, nurturing and enabling, with special technology and policy tools, specifically the *hubs* in the dense inner core – the critical high degree users and articles – that hold the key to the culture of the online community, and consequently, to its ongoing survival.

References

- Axelrod, R. (1986) “An evolutionary approach to norms”. *The American Political Science Review*, 80(4):1095–1111
- Barabasi, A. and Reka, A.. (1999) “Emergence of scaling in random networks”. *Science*, 286(5439):509–512.
- Billari, F., Prskawetz, A., and Furnkranz, J. (2003) “On the cultural evolution of age-at-marriage norms”. In Billari, F. and Prskawetz, A. editors, *Agent-Based Computational Demography: Using Simulation to Improve Our Understanding of Demographic Behaviour*, *Contributions to Economics*, pages 139–157. Physica-Verlag.
- Borgatti, S. and Pacey, C., (2003) “The network paradigm in organizational research: A review and typology”. *Journal of Management*, 29: 991–1013.
- Carley, K. (1991) “A theory of group stability”. *American Sociological Review*, 56(3): 331–354.
- Chesney, T. (2006) “An empirical examination of Wikipedia’s credibility”. *First Monday*, Volume 11
- CliffsNotes.com. “Cultural Norms”. Accessed 12 Mar 2009
<<http://www.cliffsnotes.com/WileyCDA/CliffsReviewTopic/topicArticleId-26957,articleId-26853.html>>.
- Cothrel, J. and Johnston, B. (2007) “Online Community Return on Investment: State of the Practice 2007.” Presented at: *Online Community Business Forum* in Sonoma, California.
- Dawkins, R. (1976) *The Selfish Gene*, Oxford University Press, ISBN 0-19-286092-5
- Drucker, Peter (1999) *Management Challenges of the 21st Century*. New York: HarperBusiness.
- Epstein, J. (2001) ”Learning to be Thoughtless: Social Norms and Individual Computation”, *Computational Economics*, 18: 9-24.
- Giele, Z. and Elder, G. (1998) “Methods of Life Course Research: Qualitative and Quantitative Approaches”, Published by SAGE, ISBN 0761914374
- Granovetter, M. (1985) “Economic action and social structure: The problem of embeddedness”. *American Journal of Sociology*, 91(3):481–510
- Guimerà, R., Uzzi, B., Spiro, J., Amaral, L. (2005) “Team Assembly Mechanisms Determine Collaboration Network Structure and Team Performance”, *Science*.
- Heckhausen J., (1999), “Developmental Regulation in Adulthood. Age-Normative and Sociostructural Constraints as Adaptive Challenges”, Cambridge, Cambridge University Press.
- Kilduff, M., and Corley, K. (2000). “Organizational culture from a network perspective”. In Ashkanasy, N., Wilderom, C., and Peterson, M. (Eds.), *Handbook of organizational culture & climate*: 211–221. Thousand Oaks, CA: Sage
- Kittur, A., Chi, E., Pendleton, B., Suh, B., and Mytkowicz, T. (2007) “Power of the few vs. wisdom of the crowd: Wikipedia and the rise of the bourgeoisie”. *Annual ACM Conference on Human Factors in Computing Systems*
- Kolbitsch, J., and Maurer, H., (2006), “Community Building around Encyclopaedic Knowledge”, *Journal of Computing and Information Technology*, Volume 14, Issue 3, pp. 175-190
- Molina, J. (1995). “Analysis of networks and organizational culture: A methodological proposal.” *Revista Espanola de Investigaciones Sociologicas*, 71–72: 249–263.
- Newman, M. (2002) “Assortative Mixing in Networks”. *Physical Review Letters*
- Reagle, J. (2007). “Do as I do : Authorial Leadership in Wikipedia”. *Proceedings of the 2007 international symposium on Wikis*
- Rheingold., H. (1993) “The Virtual Community: Finding Connection in a Computerized World”, Addison-Wesley Longman Publishing Co., Inc. Boston, MA, USA
- Squire, K., and Johnson, C. (2000). “Supporting distributed communities of practice with interactive television”. *Educational Technology Research and Development*, 48 (1), 23–43.