Group Related Phenomena in Wikipedia Edits (v0.1)

M. Burgess and R.I.M. Dunbar

September 21, 2023

Abstract

We report on the emergence of predictable group sizes for content editing on Wikipedia, and use these to propose an explanation of group dynamics. The data show an emergent coherence in the sizes of groups formed transiently to edit the content of subject texts, with two peaks averaging at around $N = 8$ for the size corresponding to maximal contention, and peaking at around $N = 4$ over the whole distribution, with a long tail. The numbers are consistent with Dunbar's conversational group predictions, as well as general group hierarchy. We propose an explanation building on the Promise Theory of trust. and offer a scaling law that we hypothesize may apply for all group distributions based on seeded attraction. Some caveats may apply for direct comparison with the hierarchy of social group sizes owing to the activity of bots. The results have some implications for the governance of the Wikipedia commons and the security of the platform and other similar platforms and associations.

Contents

1 Introduction

Wikipedia is one of the few online platforms that openly shares the histories of editing interactions on its pages. It provides an opportunity for studying social phenomena involving both humans and machine processes. While examining these data as part of a project about the role of trust in cybernetic systems [1], several features of group dynamics were observed, posing questions about group formation from the unusual privilege of an essentially complete history of interactions.

The data show an emergent coherence in the sizes of transient groups formed to edit the content of pages on Wikipedia, averaging at around $N = 8$ and peaking in frequency at around $N = 4$, with a long tail. We obtain a simple formula for the probability distribution of transient group sizes observed during editing and propose an explanation.

2 Method

Wikipedia has a wealth of pages bearing content on a wide variety of topics, which means that it samples contributions from a wide variety of individuals with different motivations. We can assume that this diversity of topics says something about the high level of entropy of the writing pool, which in turn allows us to study the universal aspects of group phenomena, i.e. those that depend little on context or identity. Not all editors are interested in every topic, but the existence of so many pages on so many subjects, but if the pool is large enough, one can expect a relatively unbiased attraction to all topics.

By selecting topic pages and their editing histories at random we have been able to examine basic statistical characterizations across the large reservoir of intentional behaviours exhibited by online users. Certain pages were also cherry picked as controls that one might expect to be special outliers due to their subject matter, though these did not naturally separate into a category of their own in any obvious manner.

Analysis of timelines is somewhat time consuming so sampling was performed in batches. After some initial tests to determine the scale for statistical stability, the number of samples was limited to around 800 topics involving around 200,000 separate online contributor identities. The data were then checked for stability by recombining the pages into different subsets to discover stability under sampling. The results, while noisy, exhibit a robust statistical stability for samples over a hundred topics. Our final results are based on the full merged set of data for completeness.

- We use the high level of entropy of the pages and contributors to invoke a Monte Carlo approach, sampling random groups of pages and averaging the results. Analyzing the pages is time consuming, even for a computer, so we base of results on the mutual consistency of results from independent random samples, to show invariance under sampling; then we combine derive the final results from the maximal set of 827 topics and over 260,000 users for maximum significance.

- Users come to edit a topic that they care about more or less *ad hoc*. Once one author has arrived, others will notice the changes and join in. These begin to form a group to either challenge or improve on one another's contributions, depending on the alignment of their intentions.

- Editing takes place in bursts of activity, punctuated by longer gaps. We can thus identify a series of "episodes" for every page, which we take to be causally independent. Each episode defines a group size and duration and a level of contention that one sees from the signals embedded in the historical record.

- For every topic page there is also a history page which we can read and analyze using software. The pages show evidence of user identities, which might be transient and even deliberately anonymous but which are assigned regular identifiers by the platform. We can see which edits were contested by others and the sizes, dates, and times of the edits. This allows us to construct a process timeline and to identify contention between individuals. Users can argue and even undo one another's changes, and these events can be counted.

- There is little evidence of coordination between the contributors across episodes. Coordination is essentially a stigmergic process via the medium of the editing produced. In part, it is also a confrontational process, as we see from the history and discussion pages.

- Contention between users can be identified as when one user undoes the contributions of another, rather than adding to them. From even a cursory inspection of results, it's clear that contention between users is a significant issue in editing on all pages, so special focus was attached to the level of overlap between edits and on cases where one user undid the changes made by another.

- We also note that, in excess of 20% of all edits were made by automated "bots", some of which are designed to patrol changes for such activity (as well as trivial issues like spelling corrections), so the intentional aspects of behaviour are to some degree invoked by proxy.

It's trivial to collect basic quantitative measures like the length of text and the number of different user identities from the platform. More interesting are the features which distinguish selective engagement from users merely bumping into one another. One expects the subject matter, language, and semantic content to play a strong role here in attracting users to a specific page. We can show that—over the total sample—subject matter acts mainly as a seed for attracting a particular sample of users. However, this selection eventually gets eliminated as a variable, contributing mainly to noise, as the pool of editors is sufficiently statistically similar to see no particular correlation with behaviour and subject matter. There is as much contention for cabbages or mathematics as for politics and celebrity. This is presumably because the reasons for contention and criticism are somewhat similar across all subject areas, just as typical errors of production are common to all forms of manufacturing.

The technique for transmuting semantics into countable results has been described previously in [2,3]. Briefly, the method is able to select fragments of language somewhat analogously a bioinformatic analysis, by fragmenting symbols into n-grams. The relative frequency spectrum of n-grams allows one to identify significant content independently of language. The approach has been tested on European and Chinese texts with similar results, but the final sample of pages was limited to mainly English language pages.

3 Results

Figure 1 shows how the number of users involved per editing episode varies with the current snapshot length of the main article. The article length may be taken as a proxy for the total time of the page's history on a human scale. We see that rate of user comings and goings users decays almost imperceptibly with length, suggesting a slow stabilization of topics. However, this is small compared to the level of activity maintained over time. In essence, we see little evidence of pages ever being finished.

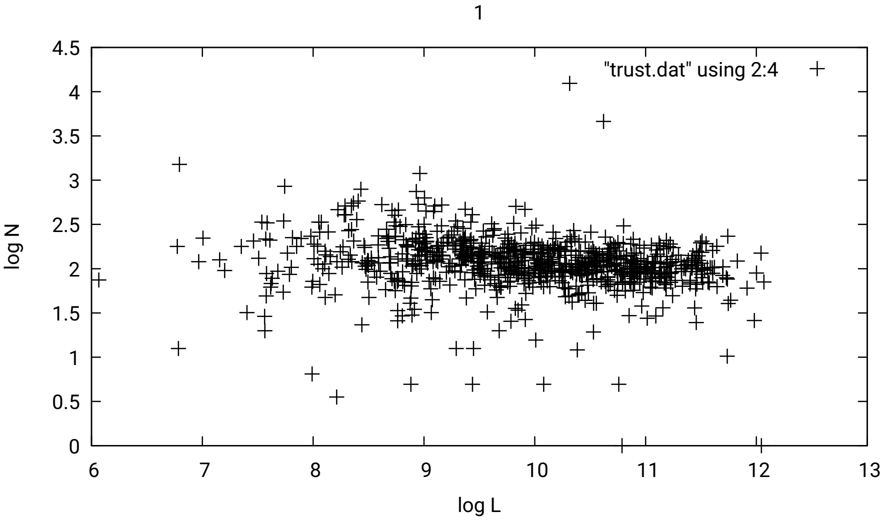

Figure 1: Relationship between the average number of users contributing per episode as a function of article length, $\langle N \rangle (L)$ is also noisily constant or decays in a weakly exponential manner, suggesting that there is a more or less constant stream of new battles to be fought or issues to be solved. The log plots show a weak tendency to exponential behaviour as one expected from a probabilistic arrival process, however the restricted number for N evident in all the graphs makes this almost insignificant.

All pages appear to be subject to continuous contentious editing. The rate of stabilization with length or time is extremely slow, which one sees by using the length of the article as a proxy to total time elapsed since creation. The bursty nature suggests that causal behaviour is limited to individual episodes, so we treat each episodic burst as a separate event.

In terms of longitudinal changes over time, the first thing one notices is that the editing of a document proceeds in episodic bursts of somewhat similar duration. These are measured by computing average durations and identifying the punctuated bursts by the gaps between them. The lifetime over which a group edits is closely bounded in duration, although the bounds are very noisy. Figure 2 shows the existence of at least two distinct classes of episode duration evident in widely separated clusters. These are not artefacts of the statistics; the same features recur in different sample sets, but most of the activity is in the lowest band. It remains to be studied whether there is significant turnover in the groups with very long durations. This would tend to make sense given that the size of the group remains statistically constant over even exponentially longer interactions.

From the episodic nature of edits, we infer that users are attracted to an editing episode by the initiation of changes by an initial user seed. We'll discuss why this may happen elsewhere [4]. The number of users in a typical episodic event is noticeably clustered. We see this in a few measures, the most interesting of which is the plot of contentious changes versus size of users in an episodic cluster.

An average group size associated with contention is seen in several data plots, lying at around $N = 8$. We take this approximate scale to be the group size at which contention reaches its maximum limit (there's a variability of between 7.75-8.2 depending on how we calculate and sample the average). We add, however, that upwards of 20% of all users on the platform were approved bots working by proxy on behalf of both the platform and others. Bots perform all kinds of administrative functions like labelling, fixing citations and spelling, etc.

Wherever users come together, they disagree. One can measure contention in a number of ways. Numerically we can see from text semantics signals of disagreement, and the size of changes. The most obvious signals for this are users who undo one another's changes. If we only count such events, without

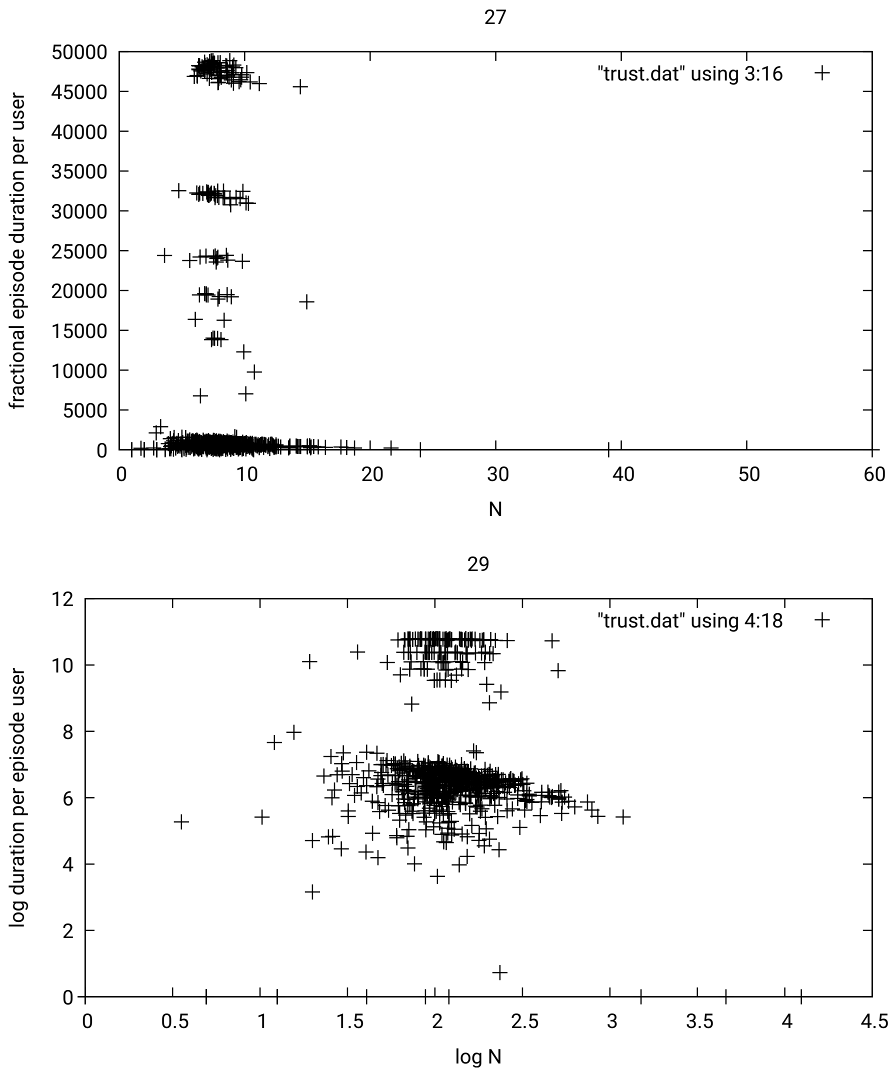

Figure 2: Duration per episode versus N. This is the average fractional time spent by a single user over an episode. The duration is measured in days. The graph axes all show clustering in a region, rather than a line relationship suggesting that the tendency to continue interacting (fighting or discussing) decays weakly and exponentially with the number of users, but is almost constant. The number of users itself doesn't vary much in episodes, so the cluster is narrow.

measuring their intensity, we find a fairly constant rate of contentious edits. The contention itself is also bounded, presumably by the size of the group, into a region. Figure 3 shows an enlargement of the non-empty region of the plot of contention per event as a function of users per episode. This shows that, although contention varies a bit between different topics, the size of the group and the level of contention is bounded to a relatively predictable region. This tells us both that contention is intrinsic to users and that users form groups of predictable sizes.

These and other measures have significance for several areas of study, including online platform security. However, we focus here on the dynamics of groups as a phenomenon in order to relate the study to previous discussions of human group formation. Already of interest is the apparent innateness for contention at around $N = 8$. The reason for this remains to be explained, and we'll return to this elsewhere [4]. For now, we focus on the empirical evidence.

We emphasize here that the cluster of group sizes is all predominantly within the range $N = [0, 15]$, centred on $N = 8$. There are no other clusters for the process. This suggests that there is single causal explanation for this cluster size.

The most striking summary of the evidence of group size coherence may be seen in the summarial frequency plot (figure 4). This suggests evidence of an underlying determinism, albeit on a statistical

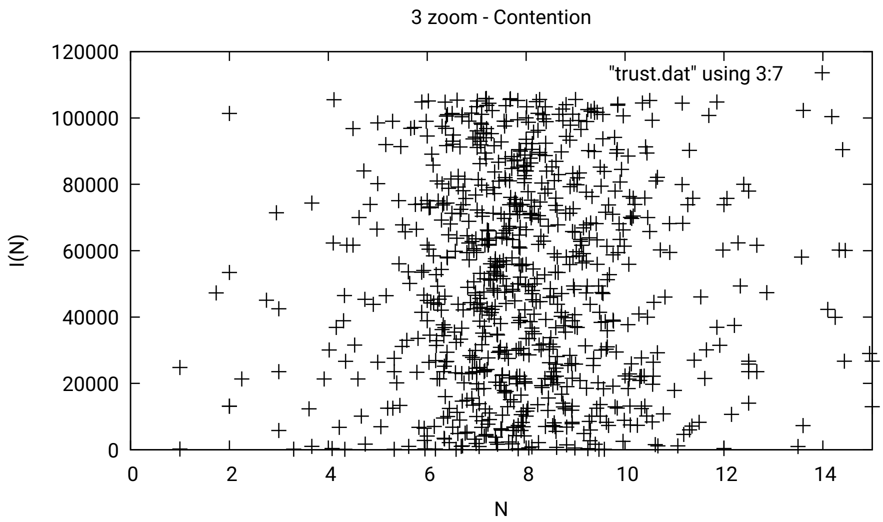

Figure 3: A zoomed view of the contention $I(N)$ and mistrust, both peak around approximately $N = 8$.

level. The figure shows crosses for actual data and a dotted line for a theoretical fit.

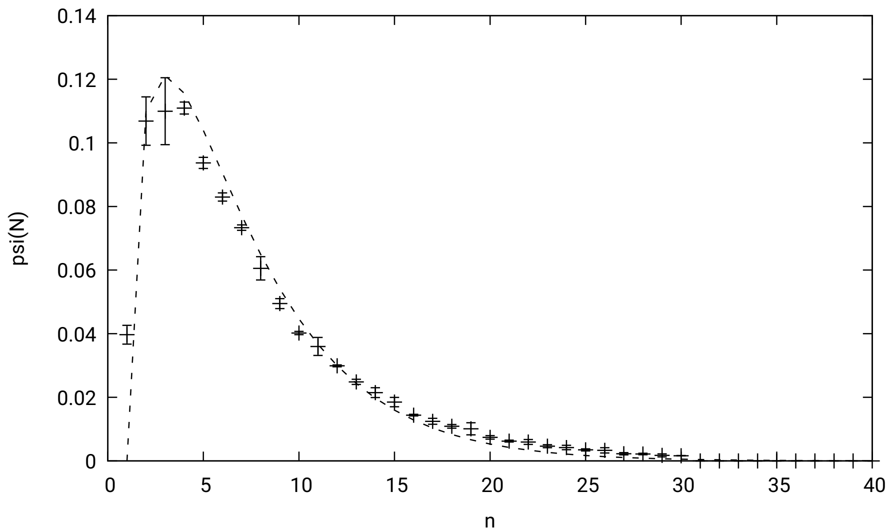

Figure 4: The normalized frequency spectrum of group sizes indicates the probability of finding an episodic group of size N. We note that the peak near $N = 4$ is well below the group size that maximizes contention at $N = 8.2$. The stippled line shows a theoretical fit for the curve derived from Promise Theory and dimensional considerations. Crosses gauge estimated error uncertainty.

In figure 4, we see the spectrum of group sizes across all episodes. It's remarkably free of the noisy artefacts from specific measurements, so we expect it to be quite robust. Error bars are no larger than the crosses. It has a classic decaying exponential tail typical of entropic processes. More significantly, it also has a non-exponential growth feature for small N, where we expect the interesting dynamics to be.

Since we know that users come to join episodic groups coincidentally by randomly searching and only later following, the story suggested by this spectrum is that users accrete into groups gradually, based on topic and activity, but on reaching a certain critical size they fizzle out. Initially the probability of adding new users to a group begins to falter around a maximum of $N = 4$, which last the longest. By the stage at which contention is maximal at $N = 8$ the probability of new attachments is already significantly reduced. Since we know the group episodes terminate, from the timeline data, we conclude that the entire group disbands. Thus the groups themselves decay over their proper time durations probably as a result of general entropic diffusion. Such exponential decay is a typical statistical feature of high entropy mixing.

4 A model of group formation

We explain the results in terms of a model that combines the roles of user intent and the work effort involved in coming together. We can measure the amount of semantic "novelty" or "attention provoking content" in both the Wikipedia article texts and the supporting editing histories. We define the amount of work done by users according to the length of their text changes in alphabetic characters (for glyphs we also count the numbers of strokes to yield a similar effect), since short or easily written words are generally more common.

The amount of work is linked to the amount of attention offered by each agent, and this plays a key role in explaining the results (see below). The presence of contention with activity suggests that activity is stimulated by what we might call mistrust of the changes made by others. We can assume that few if any of the agents know one another. There is little evidence of cooperative behaviour.

In addition to the article text, which is the reason for users to come together, we can go further with the logs of change histories to see whether contributors altered each other's contributions deliberately in short succession. Over time, one expects all text might be altered in major and minor ways, but when this occurs in the manner of a duel between two users, it marks a significant semantic interaction. In such a case, we characterize the behaviour as "contentious".

The arrival and departure of users thus forms a kind of 'detailed balance' model for the statistical stability of the group spectrum, i.e. the probability of finding a group of size N. Let us assume that the set of agents affiliated with the Wikipedia platform is large and broad in its characteristics. This can be expressed by saying it has a maximum entropy. Now, suppose that a new Wikipedia topic page is started at random by any agent. Eventually another agent will find the changes and be attracted by the seed of a common interest. Indeed, a high proportion of edits is carried out by automated processes that police the platform. Coincidence doesn't imply complete alignment between the agents, and they may contend with one another at a rate proportional to the number of others in the group $N-1$. This leads to kinetic work activity proportional to $N-1$.

Suppose we assume that there is a reservoir of agents whose intentions are distributed with high entropy. Wikipedia does not advertise nor incentivize editing in any way, so the arrival of a single agent to edit a page is completely ad hoc. Using Promise Theory [5], we can assume that a topic is represented by a direction in the space of promise outcomes and that there will be a subset of the agents who are aligned with their understanding of this approximate topic.

To determine the rate of attention for the agents, we follow the guidance offered by the dimensions of work/energy for work done by a pressure or force and the resultant kinetic response. Precise details needn't be known to find the rate of temporal evolution as a velocity change from the kinetic work $Fdx = d(1/2mv^2)$. Given a fixed attraction pressure from arriving agents, the rate of attachment would be flatten like the square root of contention "power output" the agents. From information theory, a maximum entropy conservative system follows the Boltzmann distribution $\exp(-\beta E)$, where E is an energy or work parameter. We don't actually require the number of agents to be conserved, as the number is assumed to be basically infinite. A parameter beta then represents the agents' average intolerance for contention "pressure". Choosing the only dimensionless combination of parameters $\nu \sim (N-1)/\langle N \rangle_{\overline{T}}$, where $N-1$ is the group size exerting pressure on a newcomer and $\langle N \rangle_{\overline{T}}$ is contention maximum groups scale, we obtain a probability of

$$Pr(N) = Pr(\text{attachment}) \ \textbf{AND} \ Pr(\text{dissipation}) \sim \sqrt{\nu} \times \exp(-\beta\nu) \tag{1}$$

After a short calculation [4], we find an expression for the probability of finding a group of size n:

$$\psi(\nu) = \frac{4}{\sqrt{\pi}} \frac{\nu^{\frac{1}{2}} \, e^{-\nu}}{\langle N \rangle_{\overline{T}}}, \qquad \nu = \frac{2\beta(n-1)}{\langle N \rangle_{\overline{T}}} \ , \ (n > 1). \tag{2}$$

The broad narrative implied by this result is that new pages are random events ('event one' of an episode). These events form a seed that attracts the attentions of others. in a group of total size n, $n-1$ will be stimulated by the seeding to follow the rising square root rate of kinetic attention. This corresponds to 'grooming' work in the language of [6]. As the back-pressure arising from inevitable contention in the group rises, innate limitations on agents' capacity for this work drives them away. The entropy of the total mixture ensures this is exponential on a large scale.

We need to be careful in interpreting statistical distribution laws as causative, because a statistical law does not necessarily have reverse causal implications for individual instances. A tolerance for group size derived statistically may only be a weak indicator of individual behaviour. However, there is something intriguing about the result on a number of levels. There is an 'invisible hand' style shaping of outcomes which is not quite a force, but which could be modelled as one in an effective theory. Indeed, theory suggests that there is a relationship between these outcomes and innate properties, which we discuss elsewhere [4].

The distribution relates the tendency for contention in a group to a tendency for neighbours to tolerate one another's presence. Initially, this work seems to be the very process of discussing and contending over subject matter editing changes. Later this progress seems overpowered by the tendency for contention to overcome the perceived benefit. This interpretation makes sense because the contention maximum always lies at larger size than the most common maximum size. In other words, contention is still growing when the group starts to falter.

5 Universality of the distribution

Before moving on, it's tantalizing to speculate on the wider behaviour of the curve in figure 4. Implicit in this functional form is a kind of stepping stone hierarchy of scales, determined in detail by the contention cost beta. To be clear, we only have data for $\langle N \rangle_{\overline{T}} = 8$, but we can compare the consequences of the model for other values with data from elsewhere.

If we examine the relationship between the maximum of contention and the maximum equilibrium group size we see an interesting hierarchy of sizes, reminiscent of the Dunbar hierarchy (5,15,...150,500). The precise values depend on the choice of β, whose value lies close to 1 for this work, but tolerates minor adjustment to smaller values. The value that generates the usual Dunbar hierarchy of sizes is closer to 0.875 [4].

Implicit in the statistical law (2) is a prediction that there is a fission rate for groups with contentious interactions of approximately 2β for $N \gg 2\beta$ (for smaller N the integral nature of N precludes a precise analytical expression). The values in this study are consistent with the work on conversations summarized in [7], These show conversational groups at around $N = 8$ dissolving into two approximately separate groups. The emergence of a contention limit or peak level plays the role of an 'innate' characteristic for the bulk of the agents. We should not forget that 20% or more of agents are in fact software bots, with potentially infinite tolerance, however their behaviour is largely triggered by human activity. The role of attractors is a key feature, represented by the factors $(n - 1)$ in the spectrum.

In the case of the $\beta = 0.875$ series of 5,15,...150, etc, a group that could loosely contend at $N = 550$ would tend to break up into fragments of 150. A loose group of around 15 might break up into smaller groups of order 5 under pressure, and so forth. This suggests an interesting kinetic link between the dominant group scales. Again, we emphasize that these are statistical tendencies not causal rules at the network level. The network causality implicit in the formula is one based on the kinetic work done by the process supported by the local network. The precise numbers are subject to uncertainty, but the essence of the story is compatible with the scaling hierarchy of Dunbar numbers known from wider sources.

The square root growth is a consequence of the work/energy invested in attending to the ongoing process. The cognitive cost of increased familiarity is an increased processing time cost. In physics, energy is the complementary variable (and thus generator) of temporal evolution in a system. This relationship is essentially information theoretic on a statistical level. The link between intentions, promises, and trust points to this intentionality as the seed attractor in the explanation of process network dynamics [8]. The attractive 'force' during growth is not a network effect but a kinetic attention effect over these networked bonds. Attraction does not therefore imply proximity in physical spacetime, but rather in intention space, i.e. the alignment is in intention first, and in position only as a secondary consequence of attention [8].

6 Remarks

The signature of an apparently universal group dynamics is evident within the data mined from Wikipedia. We didn't set out to look for it, yet it seems to dominate the behaviour and offers a fortuitous opportunity to measure details normally unavailable to onlookers. We should underline that we have only evidence of a single grouping phenomenon with characteristic scales $n_{\max} \simeq 4$ and $\langle N \rangle_{\overline{T}} \simeq 8$; our argument for a possibly wider applicability of the mechanism is laid out more fully in [4].

The question one is left with is what is the seeding force or "invisible hand" shaping attraction and repulsion in the group processes. Is it the charismatic leader, the threat of a predator, etc? For Wikipedia, it is clearly the suspicion that someone is changing a subject other care enough about to come to its defence. The process of 'grooming' or maintaining the relationship is with the subject matter rather than the others in the group directly, and it is clearly a kinetic one: the activity is that of a standalone agent, which is therefore limited by the characteristics of the agent, some of which are innate and others a function of state.

One of the motivations for studying user behaviour in informatics is to understand the security of online platforms, like Wikipedia. These are increasingly prevalent features of our lives. In Informatics, trust is is usually handwaved away with very simple questions about identity reliability (trustworthiness), leading to misleading calls foe 'zero trust' behaviour. The challenge in using trust as a characteristic in the security of systems is that it represents something taking shape over very large numbers of events. This may well make it useless as a practical tool unless there are truly universal aspects that transfer from one case to another.

There has been a tendency to look to 'complexity science' to explain phenomena that relate to biological and social systems. We see little of that in the results here. At the scale of bulk measurement we find remarkable stability. This is an indication of a separation of timescales between interior processes involved in relationship maintenance and the exterior 'boundary' conditions of the social network. What is challenging for monitoring of systems in general is that the timescales over which significant changes may occur is must faster than the timescale over which statistical learning takes place. This is surely why attentiveness is so expensive and group sizes are constrained in a hierarchy of tradeoffs. Indeed, if we seek to compare singular interactions to an average profile of bulk behaviour, in order to weed out bad actors, we are faced with a learning challenge. This challenge is currently under scrutiny in connection with Machine Learning for Artificial Intelligence (AI). This is surely the problem for which brains have developed their phenomenal capabilities in the first place.

Acknowledgment: MB is grateful to György Buszáki for discussions about the neuroscience, and for financial support from the NLnet Foundation.

References

[1] NLnet Foundation. Trust, semantic learning, and monitoring. NLnet sponsored project website.

[2] M. Burgess. Testing the quantitative spacetime hypothesis using artificial narrative comprehension (i): Bootstrapping meaning from episodic narrative view as a feature landscape. *preprint*, 2020.

[3] M. Burgess. Testing the quantitative spacetime hypothesis using artificial narrative comprehension (ii): Establishing the geometry of invariant concepts, themes, and namespaces from narrative. *preprint*, 2020.

[4] M. Burgess and R.I.M. Dunbar. A promise theory perspective on the role of intent in group dynamics. *In preparation*, 2023.

[5] J.A. Bergstra and M. Burgess. *Promise Theory: Principles and Applications (second edition). χtAxis* Press, 2014,2019.

[6] R. Dunbar. *Grooming, Gossip, and the Evolution of Language.* Faber and Faber, London, 1996.

[7] R. Dunbar. *Friends: Understanding the Power of our Most Important Relationships.* Little, Brown, 2022.

[8] M. Burgess. Notes on trust as a causal basis for social science. *SSRN Archive, available at http://dx.doi.org/10.2139/ssrn.4252501 (DOI: 10.2139/ssrn.4252501)*, August 2022.