The Invisible Hands of Time: How Timezones Shape Online Communities

Wybo Wiersma
Oxford Internet Institute
mail@wybowiersma.net

1 Introduction

The popular view of online communities is that they transcend time and place. But as many popular views of the internet, it is partially wrong. Threads are created at a specific point in time and (mostly) ordered in chronological order. Moreover, as replies are posted the ensuing online-discussions unfold over time as well. All this makes the time at which participants arrive potentially very important. Not only will it determine which threads people will see first (positioning matters, at least for Google results), but also which comments have not already been extensively replied to.

The research reported in this essay examines this process in detail, by looking at how time and timezones affect the community structure of an international news-site called Hacker News (HN).[14] The main question examined is whether time affects the social structure of the community, and if so, in what ways. This is an important question, because if time-effects are important — and they were found to be to some extent —, then purely social interpretations of reply structures in on-line communities (common in Social Network Analysis) might have to be reconsidered, as well as wider notions of online communities, and their supposedly global, and asynchronous natures.

The paper will begin with a discussion of previous work, followed by an exposition of its theoretical background, and the expectations derived from these. Then the limits of this study will be clarified. Next the data-set, and some challenges in collection and processing will be discussed, along with some descriptive statistics. Following that, the methods used will be introduced, among them social network analysis, and three statistical tests: a permutation-test of posts in- and across time-windows, a regression of network-distances versus time-distances, and a (provisionary) t-test for the impact of daylight saving time on the network-distance between UK and US users. These methods will be explained together with their results, and the various ways in which time was found to impact the HN-community. Things are then wrapped up with a discussion of findings.

1.1 Previous work

There are several papers analysing social networks in thread-based online communities, whether forums, usenet, mailing-lists, or news-sites such as Slashdot. The surprising thing is that none of them takes time-effects at the circadian, or 24-hour scale, into consideration. They focus either on users and their roles (replyer, initiater, flame warrior), or the relationships between users, as implied by exchanged messages (reciprocity).[53,24,20,22,54,29,5,27,25,4,44,59,3,12,9,41,55,8,51]

Therefore, the net was cast wider, and several interesting papers were found. A set of papers by Kaltebrunner provide a detailed timeline of how discussions unfold on Slashdot.[21,23,43,49,43] Another work noted variations in the number of active players in Multi User Dungeons (MUDs) at different times of day.[6] A daily cycle in usenet traffic was reported as well, and yet another paper showed that (in the early days) usenet threads extended over multiple days.[39,47] Then there is a paper which introduces a way of displaying e-mail interactions over time, as occurring on concentric circles.[28]. Finally, daily cycles in broadband- and mobile traffic were observed.[30,58,7,46]

Though these papers sketch out the landscape and suggest the possibility, none are specifically on the effects of time differences on the social structure of thread-based communities. Wider theoretical frameworks were thus used as the main source of hypotheses, about which more now.

1.2 Theory and expectations

The approach taken in this paper, first of all, is that of analytical sociology, where the focus is on so called social mechanisms. Social mechanisms operate on the micro-level, impacting decisions made by individuals, and bring about macro-level effects through emergence.[18,17,19]. The social mechanism examined here, is determined by time. As already mentioned, people come online at certain times of day (their daily schedules or circadian rhythms). However it is expected that this, together with time-pressures that operate within threads, should make near-synchronous social interactions, more likely than asynchronous ones.

Time-pressure effects within threads are constituted by the fact that duplicate replies (or replies making a very similar point) are normally not appreciated by readers. Which makes it hard to reply (in a socially productive way) if a post has already received many replies, effectively creating a situation in which the only posts that one can reply to are those that are relatively new. In addition, early replies are read more (at least replied to more), and thus provide more visibility to their author. Whether such thread-level time-pressure effects exist, or at least can be corroborated by the data, is the first hypothesis tested in this paper (see section 3.1).

The second hypothesis is that what appears social (which relationships people form), and is normally ascribed to factors such as reciprocation, friendships, and other directly social factors, is in fact dictated (to some extent) by differing circadian rhytms and the thread-time-pressure effect: this by making it much more likely that people with overlaps in the time they come online, reply to one another's posts (this is tested in sections 3.3 and 3.4). The theory that informs this hypothesis, is Structuration Theory (not to be confused with structuralism), which states that besides individual agency, societal, and other structures, shape social behaviour.[11,38]

Affirmative findings for both hypotheses could be used to place doubt on the social nature of the online sphere. To the extent that online communities might best be considered imagined communities, in the meaning given to this by Benedict Anderson.[1] Anderson uses the term to describe modern nations as 'communities' imagined by the people perceiving themselves as part of them, rather than as real ones, of people that know one-another on a personal basis, and are interacting on a regular basis. Finally, finding that time has an influence, might indicate that the Global Village is partitioned up by timezones.

1.3 Limits

When time is discussed in this paper, however, the absolute time at which people normally post, in GMT, will be considered, not their timezone, or things such as whether people are online in the morning or evening. This because, given that timezones differ across the globe, an English night-owl and an Indian early bird may very well sit down on the same discussion-tree every day, while a nocturnal American may never reply directly to his diurnal flat-mate frequenting the same site.

A related limit is that, as we don't have complete information on the locations or timezones of HN users, in this paper no clear distinctions will be made between time-differences and spatial distance. Given that communities traditionally have been defined by spatial closeness, this leaves room for sceptics to argue that any time-based findings are solely based on differences in spatial distance. We mostly concede to this (though in section 3.4 counter-evidence is provided). A similar limit is that the data used here is not tagged for topics, and thus if similar topics were consistently posted at similar times, then time-effects might be an artefact of differences in interests.

It should also be kept in mind that to the extent time effects are found, they do not exclude social effects, but set their stage, at most. And this is no different in the offline realm, where, for example, it is ultimately not possible to form a reciprocated friendship with somebody from the 18th century. Yet on the other hand, while living close together in time and space may afford a community to form, it is never sufficient cause (as the lack of community in many suburbs makes clear).

Another caveat is that the relative impact of structural and social factors (even if beyond doubt), will heavily depend on the specific online community under consideration. Small, localized communities centred around personal support, will, for example, be likely to display much stronger social effects, than large communities, focusing on technical topics. Also, this paper only studies one such community.

2 Data and descriptives

2.1 Sampling

As this is the first study on circadian time-effects on web-forums, and it is a relatively labour intensive process to create/pre-process data-sets, the sampling strategy was not based on a random selection of sites. Instead we picked Hacker News (HN) as a case study to explore whether time-effects play a role at least in some communities (this one).

HN is comparable to Slashdot, in that people can post links to news stories, and comment on them.[43,14] It has recently surpassed Slashdot in visitor-numbers, and is quite large, with 100,000 unique visitors per day, most of whom are lurkers (see fig 1).[13] It is ran by Paul Graham, an internet millionaire, who also runs Y-Combinator, the most competitive startup investment/incubator program in the world.

What makes HN especially suitable for finding time-effects, is that its main user-base, the startup- & web-development-scene, is very international. A glimpse of this can be gotten from the following map, which shows 1500 active users that volunteered to mark themselves out on a map (see figure 2).[31]. This global nature makes it more likely for time-effects to play out and be detected in this data-set.

2.2 Processing

A set of Ruby scripts were written to gather and process the data, which was done in stages. An initial script fetched the threads and user-pages. Another parsed them on the local machine using the Nokogiri HTML scraping library.[35] Further scripts were created for generating and saving networks, checking the integrity of the data, and for calculating statistics. Other tools were employed, such Stata, the R igraph package (for network metrics), and NodeXL, Pajek, and especially Gephi, for exploring and visualizing networks.[45,48,34,40,10] The custom scripts totalled 3,000 lines, which according to the Ohloh code-analysis site is over 4 man-months worth of programming.[36] The code is open sourced, and available on Github.[57] Apart from enabling a thorough analysis, some of this code was needed to overcome a special challenge with the HN data.

The challenge was that HN does not provide time-stamps with posts and threads, but rather phrases such as 'X minutes ago', which complicated getting a precise handle on the time dimension. This is especially true for material older than an hour, as minutes are no longer reported after an hour (nor are hours re-

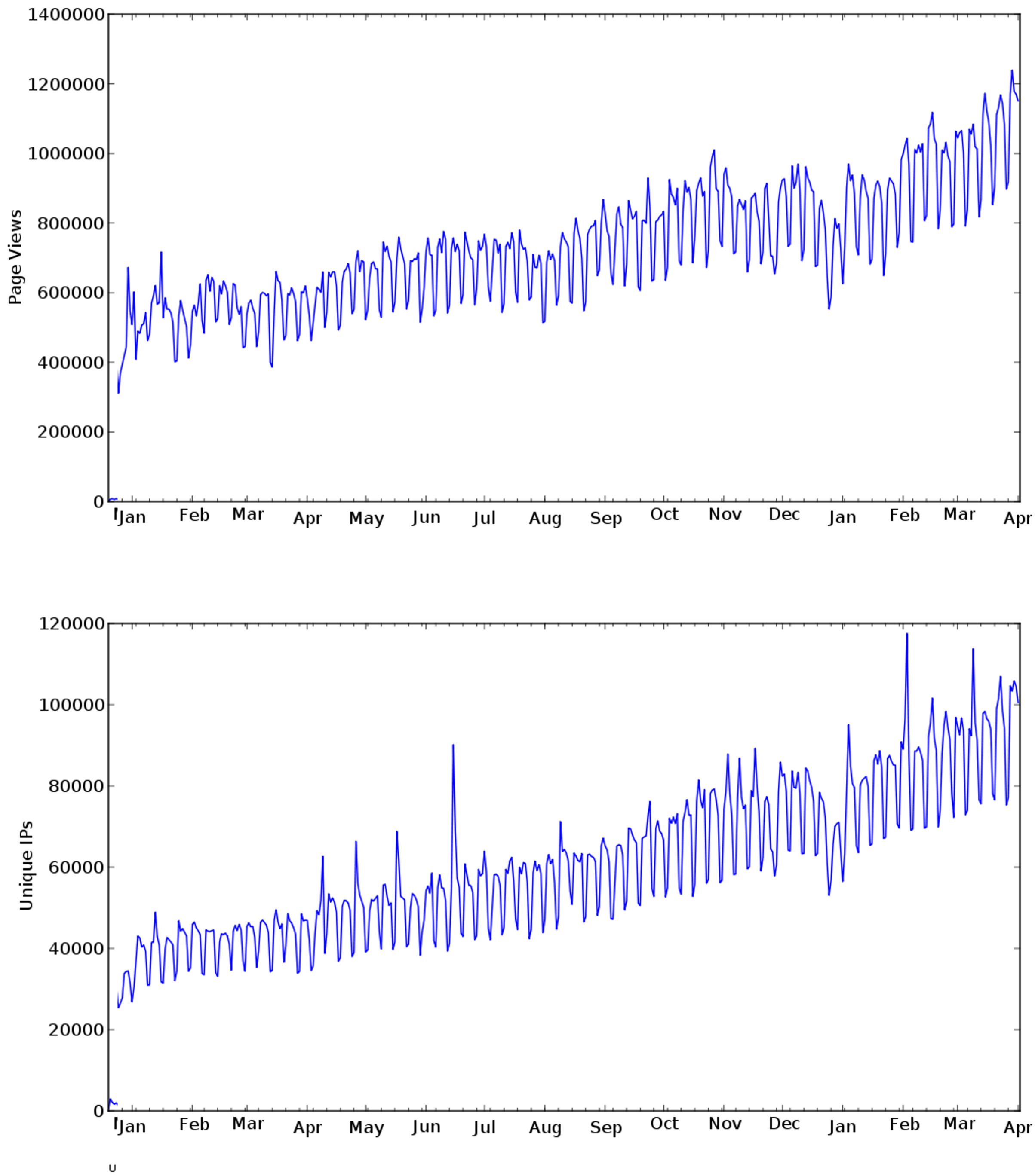

Figure 1: *HN traffic since founding (thanks to Paul Graham).*

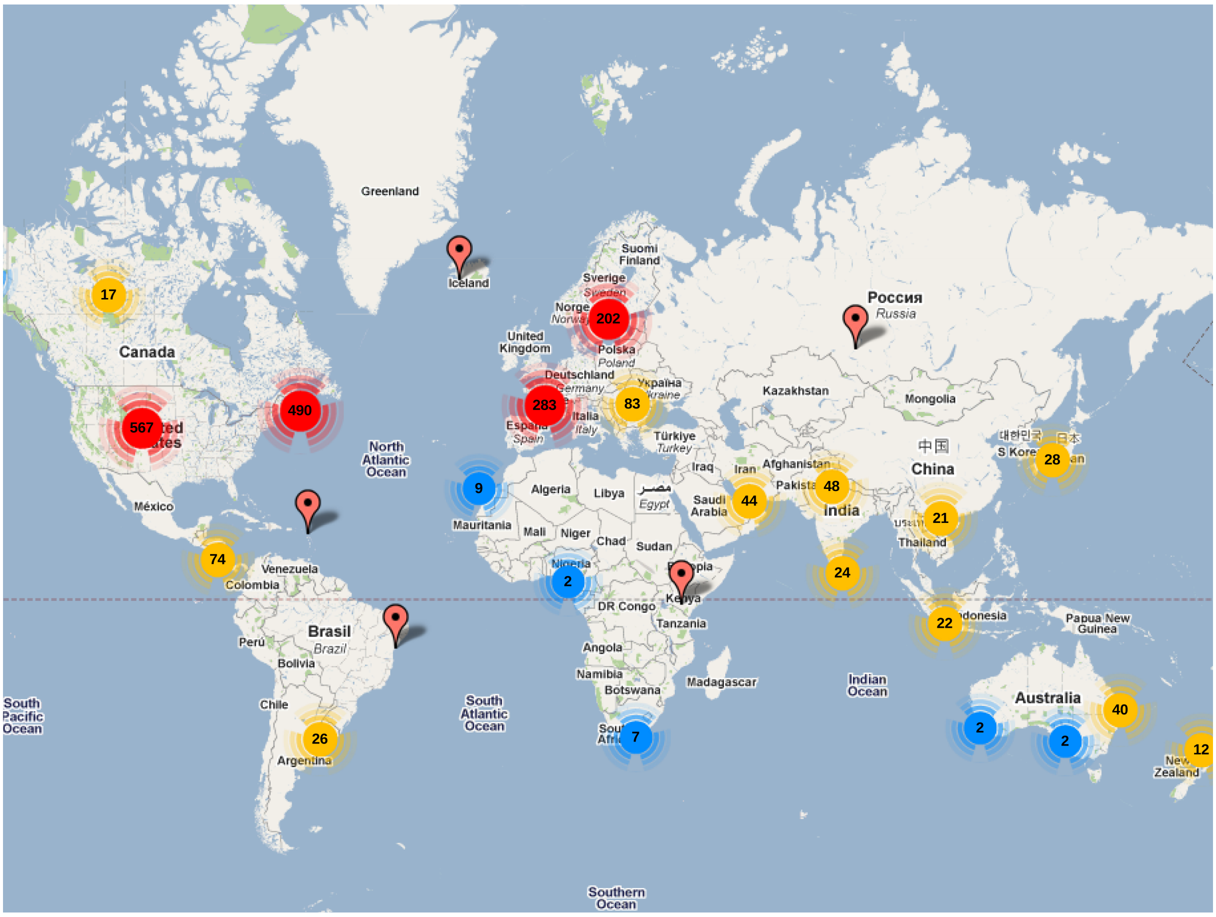

Figure 2: *Locations of HN users (thanks to John Marbach, and Charles Pick).*

ported after a day). Luckily enough, all items (posts and threads) on the site are assigned an unique ID, which increases predictably with time. This allowed us to deduce the time at which items were posted from the order of their IDs. By knowing of two items that one had been posted six minutes ago, while the other was posted four minutes ago, it could be deduced that a third item, with an ID in between these, must have been posted before the latter, and after the former, so approximately five minutes ago.

Having this knowledge made it possible to reconstruct the time at which all comments were posted, without having to sample every thread every hour in the hope of recording the time of any new comments. By another strike of luck, the site has a 'comments' page, which listed all new comments. This page was harvested every 10 minutes (from 2 separate machines, in case one would go offline), producing a sample of IDs and their timing dense enough not to leave gaps larger than two minutes. Threads thus had to be harvested only once, 24 hours after they had left the front-page. A 24-hour delay was chosen because the site showed a 'more' link at the bottom of the front-page, which allowed people to read older news. Making sure the whole thread was captured seemed worth the effort.

An additional complexity was that before threads reach the front-page, they are shown on the 'new' page, from which only the best-rated posts move on to the front-page. So an X amount of time also passed before they hit the home-page. The cut-off for this depends on the ratio of votes threads received, and the time elapsed. So threads that receive many votes in a short timespan are promoted within a few minutes. The same system is used for ranking threads once they reach the front-page, and even for posts within threads. Thus on HN, posts are not strictly ordered chronologically, but as a function of time and rating ($(r-1)/(t+2)^{1.8}$, with rating r, and t being the time in hours).[15] This separates (to some extent) the effects of time and ordering, creating a clearer picture of pure time-effects, as opposed to effects depending on placement, which coincide with timeliness on sites employing pure chronological ordering.

2.3 *Descriptives*

Data was gathered for the months of February and March 2011. However, given that on March 13 daylight savings time went into effect in the US (and on the 27th in most European countries), and shifting an hour made a difference on the circadian scale (see section 3.4), for most of the analysis only the data from the 1st of February until the 12th of March was used, leaving 40 days worth of data. During this period 3,546 threads and 104,816 posts were created by 13,314 users (user-IDs). Thus on average 89 threads and 2,620 posts per day.

In line with weekly cycles in the HN visitor-statistics (figure 1), there is also quite some variation between weekdays and the weekend in terms of the number of posts created (see figure 3), with most posts being made during weekdays. Though peoples schedules would be expected to be more consistent during weekdays, and time-effects were indeed found to be slightly more pronounced during them, in the analysis presented here, weekends were not removed from the data. Because doing so would misrepresent the community, and overestimate overall time-effects.

The hourly posts are shown in the following graph (figure 4). There is a clear peak during US-daytime, but at any time of the day a sizeable number of posts are made. Individual posting patterns for 10 randomly selected users with more than 25 posts are displayed in figure 5. As can be seen there is quite some variation, and there does not seem to be a real pattern for prolific users. The same is true if they are aligned for timezones (figure 6), though, albeit vaguely, a rough peak can be discerned (morning into day). The hourly ag-

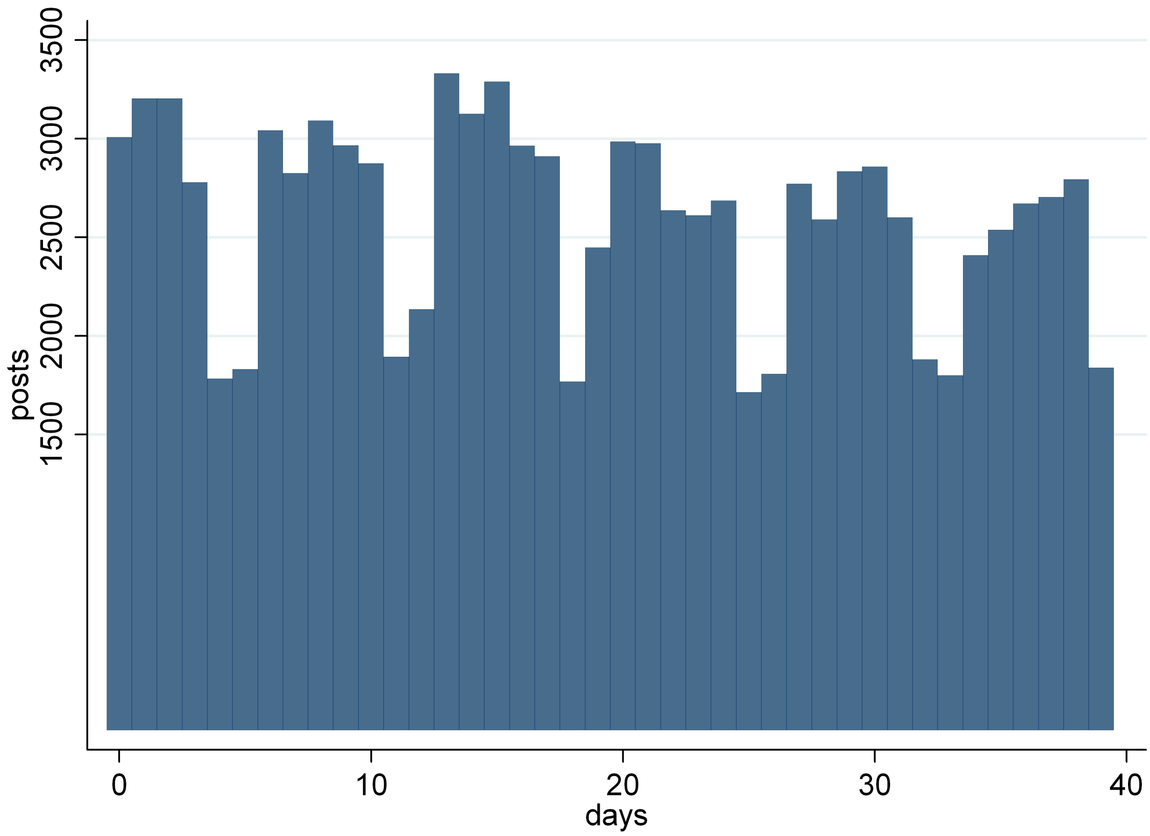

Figure 3: *Daily posts on HN from February 1 to March 12 (2011).*

gregate of all users for which timezone information could be gathered, provides a clearer pattern, as is shown in figure 7. Its shape is similar to that of UTC-based hourly posts 4, though more extreme, and shifted to the left (confirming that US-timezones are dominant on HN).

The timezone information was scraped from a sister-site to HN, to which some users registered voluntarily. The voluntarily nature of this data has undoubtedly introduced biases (their average karma is higher, and they might be more active and willing to advertise themselves). They are therefore not fully representative. However, a comparison with randomly selected users posting-patterns did not lead us to believe that their circadian rhythms are atypical. So they still give a reasonable impression of posting-patterns on HN.

As for threads, the graph with the number of threads created each hour (in UTC time, figure 8) shows a similar peak, though it is shifted slightly left-ward compared to that for posts. This meshes well with earlier findings that people creating threads are different from those that mostly reply.[3,50] Also, as on HN replies can receive replies in turn, threads can have widths greater than 2. A screenshot of a thread is shown in figure 9, and a schematic representation of two threads in figure 10. Finally, the number of posts per user, the width of threads, and the number of posts per thread, all follow power laws (no graphs shown), which means that there are many short and narrow threads, and progressively fewer long and wide ones. This in line with similar findings by Kaltebrunner (and others) for Slashdot.[21]

3 Results

3.1 Time in threads

According to the first hypothesis, time-pressure effects should be constituted at the level of threads. This will be examined now. In figure 11 the time between replies is shown. As can be seen, most replies are made within an hour or two of the post they were prompted by, with a practical maximum of six hours (though in very rare rare cases it takes longer, up to 46 hours). In a second figure (figure 12), the same effect is depicted, except that the posts are now shown on a 24-hour clock, colour-coded for the hours in UTC at which they were submitted. Edges between post-hours are reply-relationships, with arrows pointing from the prompt to the reply. As can be seen, replies overwhelmingly flow round the clock, with few more than an hour apart.

The next figure (13) shows the progress of time within three randomly selected threads (with between 20 and 25 posts). Here posts are colour-coded for the hours that passed after the thread was created (not in absolute time). As can be seen posts are made in a predictable succession. Though some sub-threads appear out of order (before earlier ones). This is because, as noted, HN raises posts (sub-threads) based on the ratio of their rating and passed time. But even then, time-effects are preserved on overall, as the aggregate for all threads shows (figure 14). Time ripples through them, top to bottom, and left to right, again suggesting a time-pressure effect.

As discussed, one of the expected causes behind the time-pressure effect was the lesser appreciation for later posts. Appreciation, both in terms of perceived quality, and number of people reading the post, was operationalized as the karma rating each post received.[42] Average ratings are shown against posting-time since thread creation in figure 15. As can be seen the karma ratings go down for later posts. Figure 14 then shows a heatmap of average karma ratings throughout all threads. Colour-codings now display the posts average rating. For this heatmap, replies for each subthread (recursively, keeping branches intact) were sorted by time, before aggregating them, to revert score-based

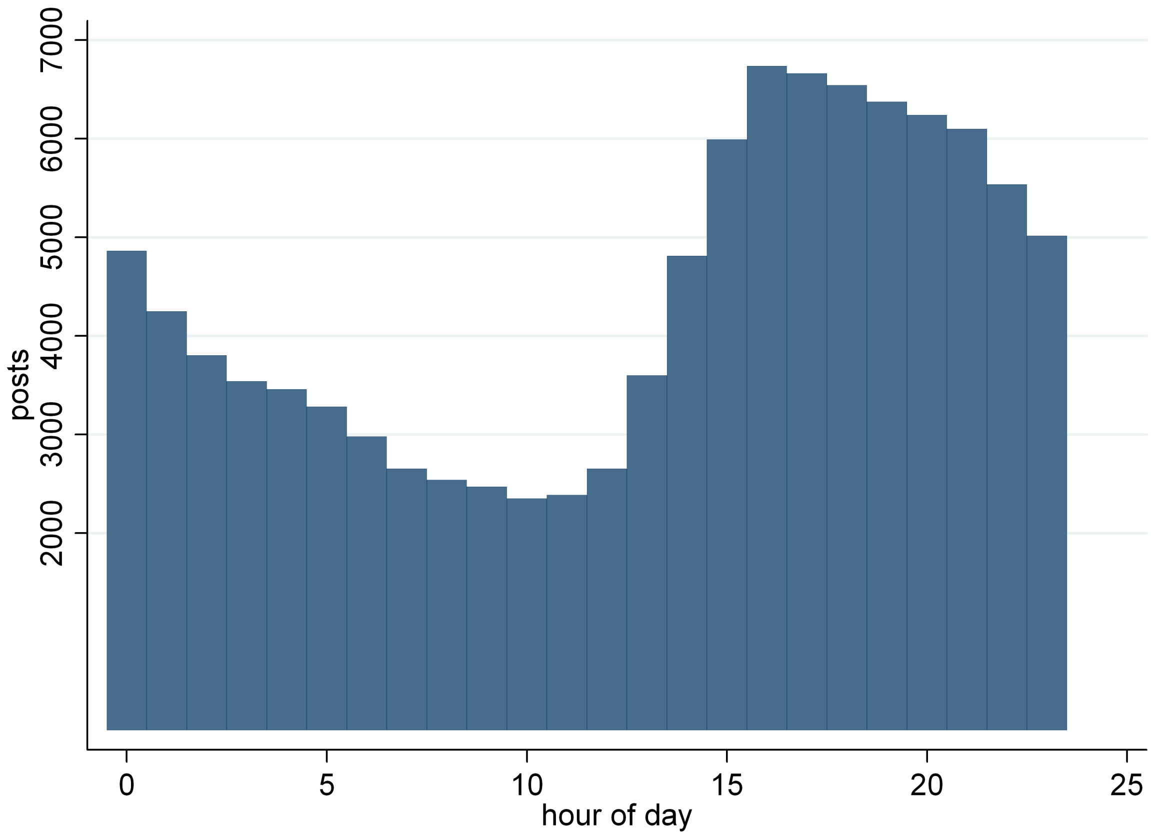

Figure 4: *Number of posts created at each hour of day (times in UTC).*

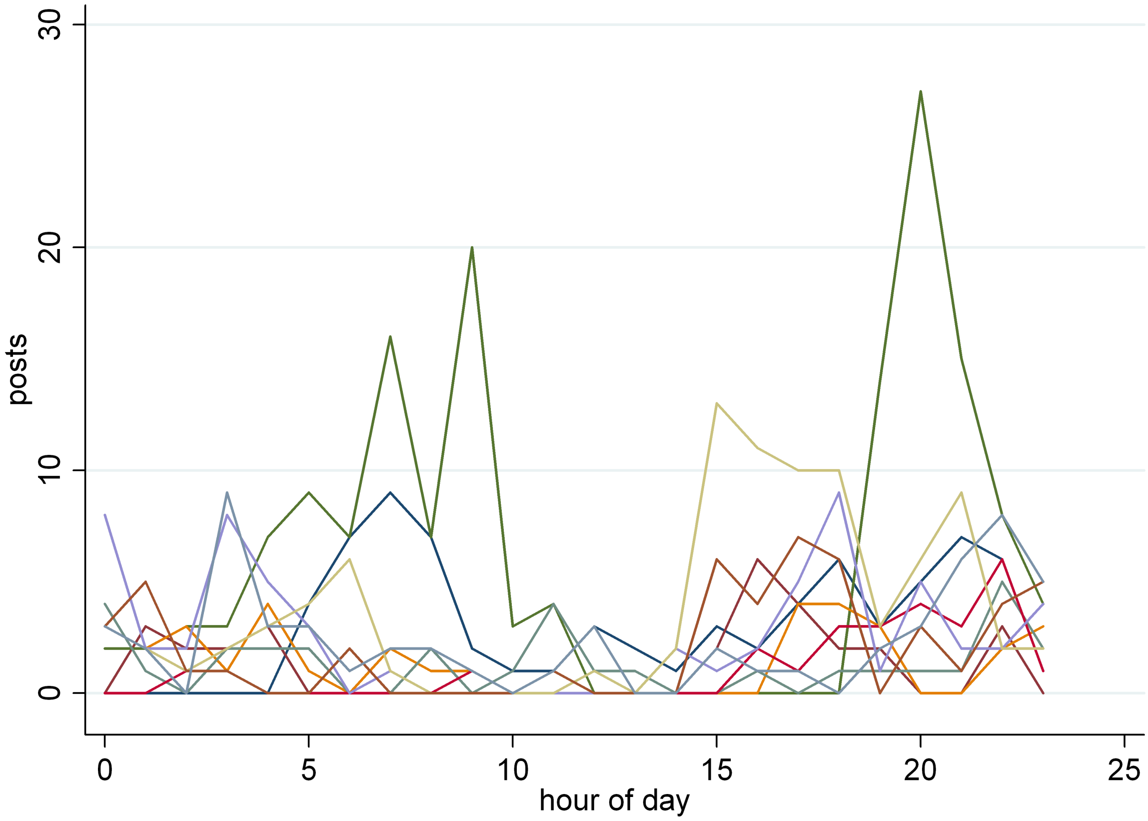

Figure 5: *Hourly posts for 10 randomly selected prolific users (> 25 posts in data-set).*

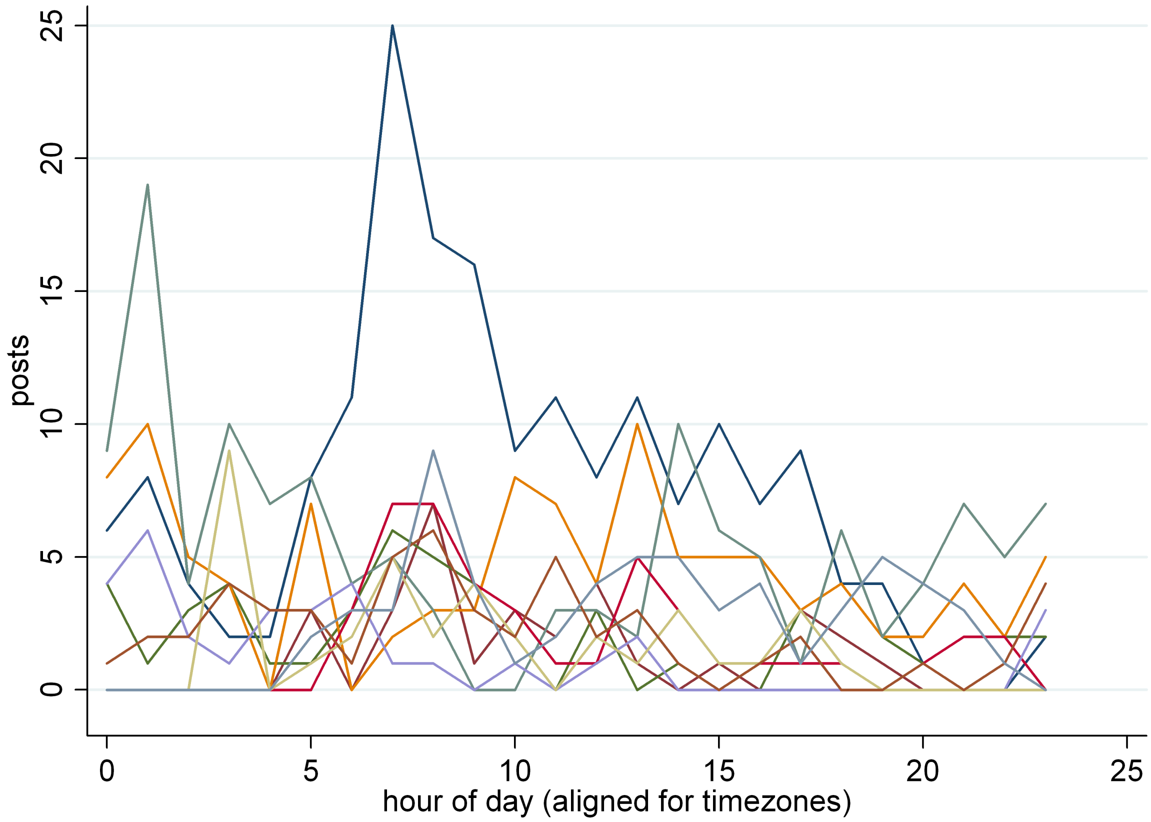

Figure 6: *Hourly posts for 10 randomly selected prolific users, aligned by their timezone.*

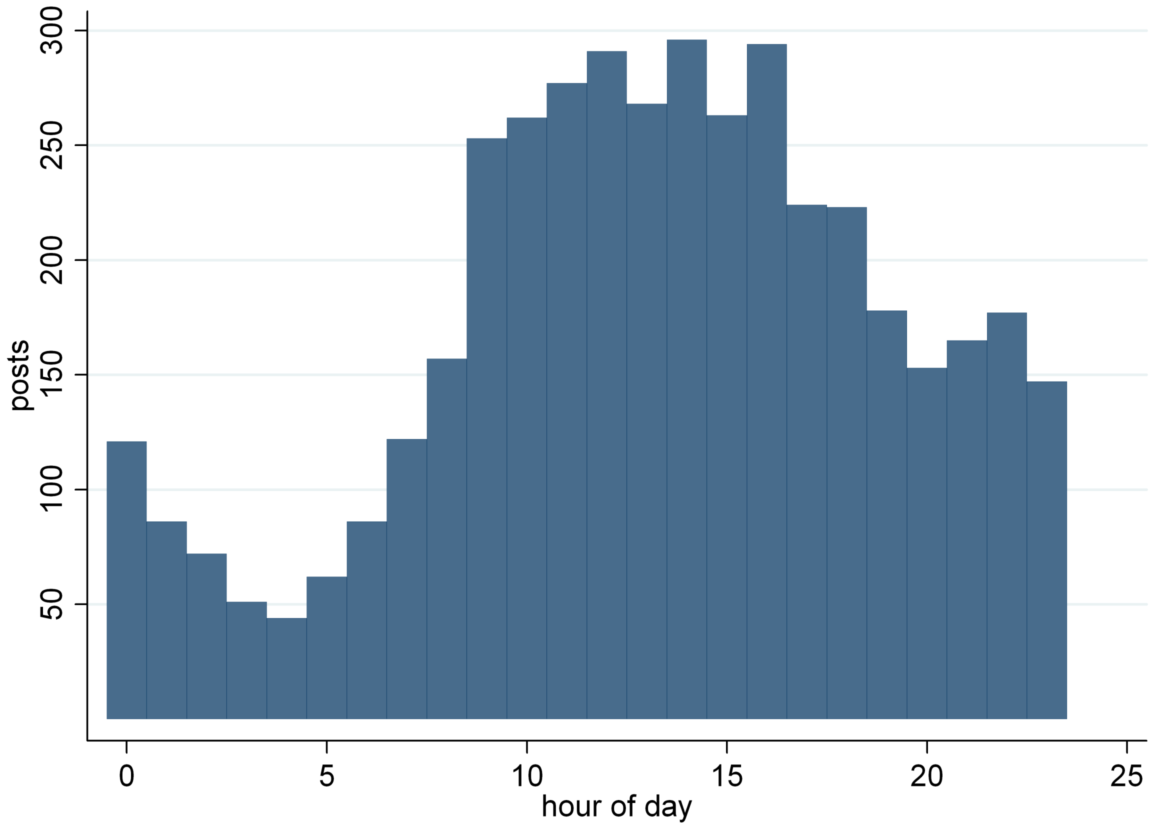

Figure 7: *Aggregate hourly posts for all users for which timezone information was gathered.*

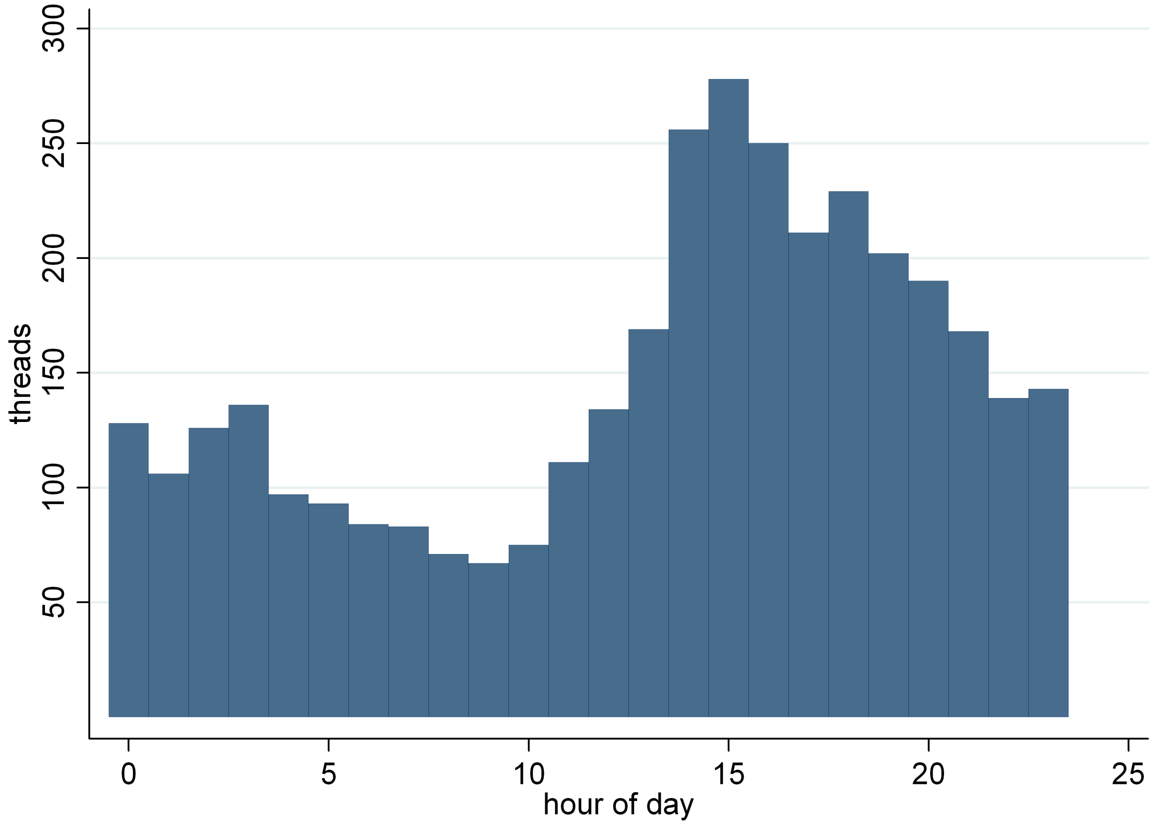

Figure 8: *Number of threads created at each hour of day (times in UTC).*


reply

▲ derleth 12 hours ago | link

What is the default mapping for Super?

reply

▲ carussell 9 hours ago | link

Generally, the Windows key.

reply

▲ xcallemjudasx 18 hours ago | link

Isn't having all of these different key bindings pointless when GnomeDO can bring them all up?

reply

▲ fingerprinter 18 hours ago | link

Sure, it might be if GnomeDo was installed by default and tweaked to do the same thing. But since it isn't, I would hardly call it "pointless", but rather, awesome.

reply

▲ technomancy 16 hours ago | link

It seems like that's a big theme with the latest Unity changes. Huge improvements over the default desktop, but nobody *I* know was using the default desktop; they all heavily modified it with devilspie, xmonad, and friends, which are far beyond what Unity offers.

So while it seems like a good development in the abstract, it's hard for me to get excited about it personally.

reply

▲ gcr 13 hours ago | link

One reason I used wmii instead of the default is because the default stunk. Unity is raising the bar on the default desktop now; I'll definitely give it another look.

reply

▲ nakkiel 1 hour ago | link

I second this. I read some information on Unity and it all sounded so common to me that I thought WMII and the likes actually had an impact on how people making window manager think about window managers.

I'll stick with WMII though as it never gets in my way.

reply

▲ hasenj 8 hours ago | link

> but nobody I know was using the default desktop

Because the default desktop setup sucks.

But not anymore.

Unity makes things gnome-do and docky *obsolete*.

And please don't compare it to hacker-centric window managers.

reply

▲ curtin 20 hours ago | link

Those keybindings look awesome and make me want to upgrade. Guess I will try it out in a VM first though.

reply

Figure 9: *Screenshot of (part of) a thread on HN.*

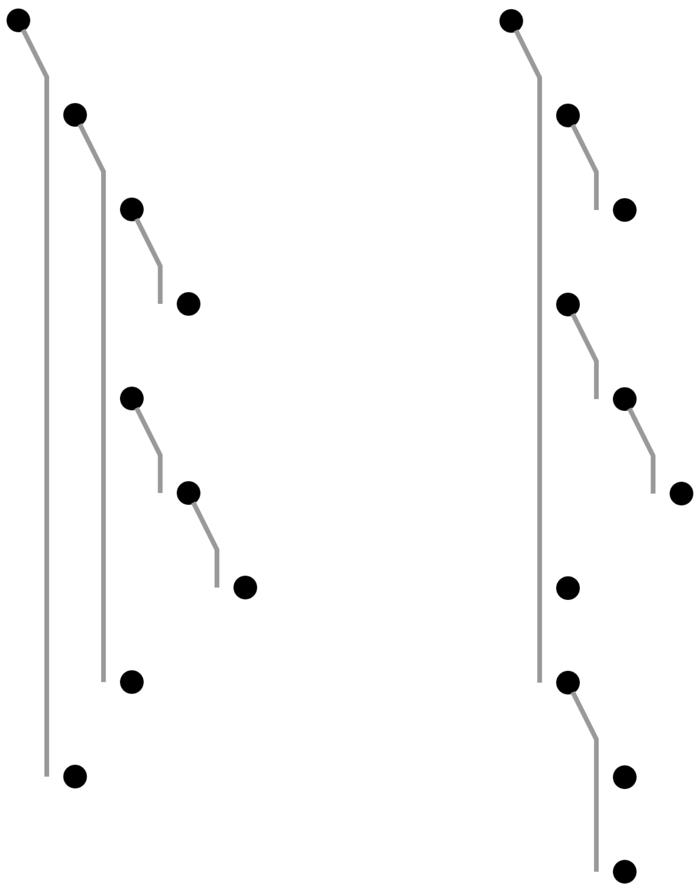

Figure 10: *Schematic representation of two randomly selected small (8-12 posts) threads.*

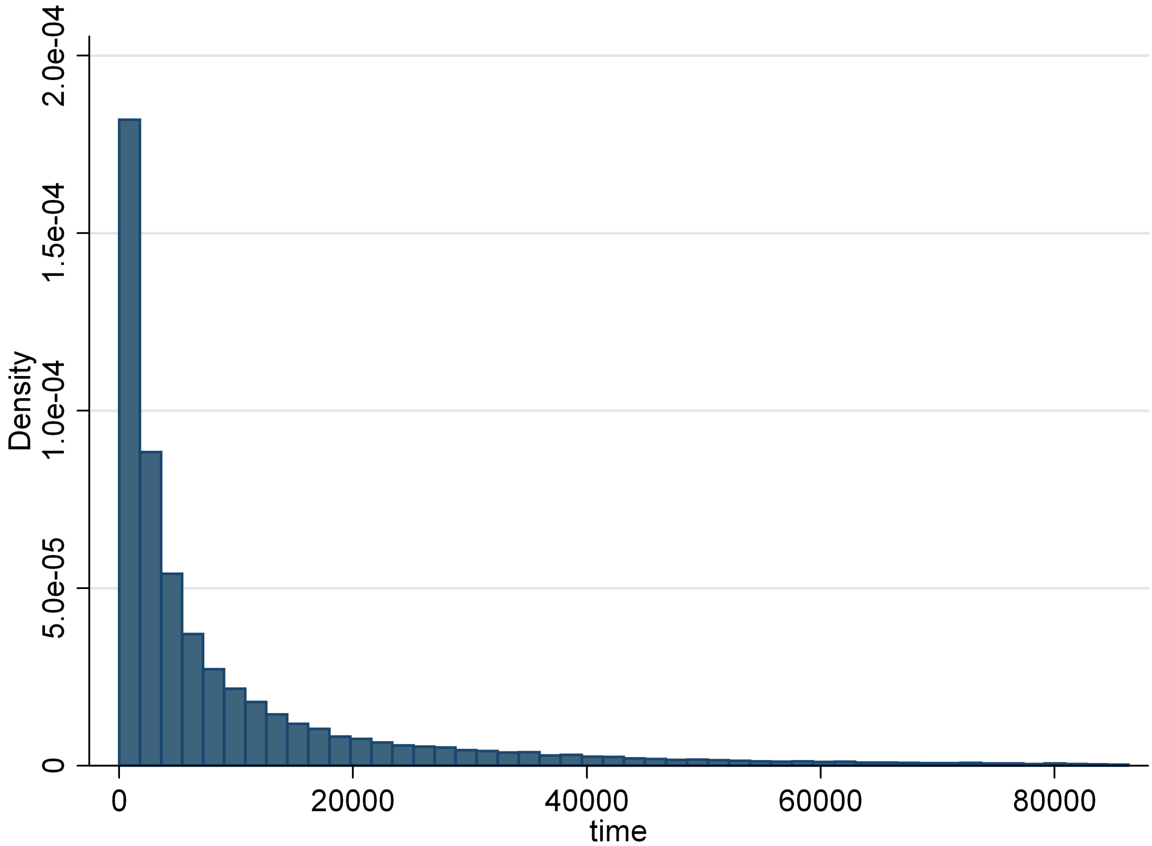

Figure 11: *Time in hours between replies and the post they were prompted by. Total numbers shown.*

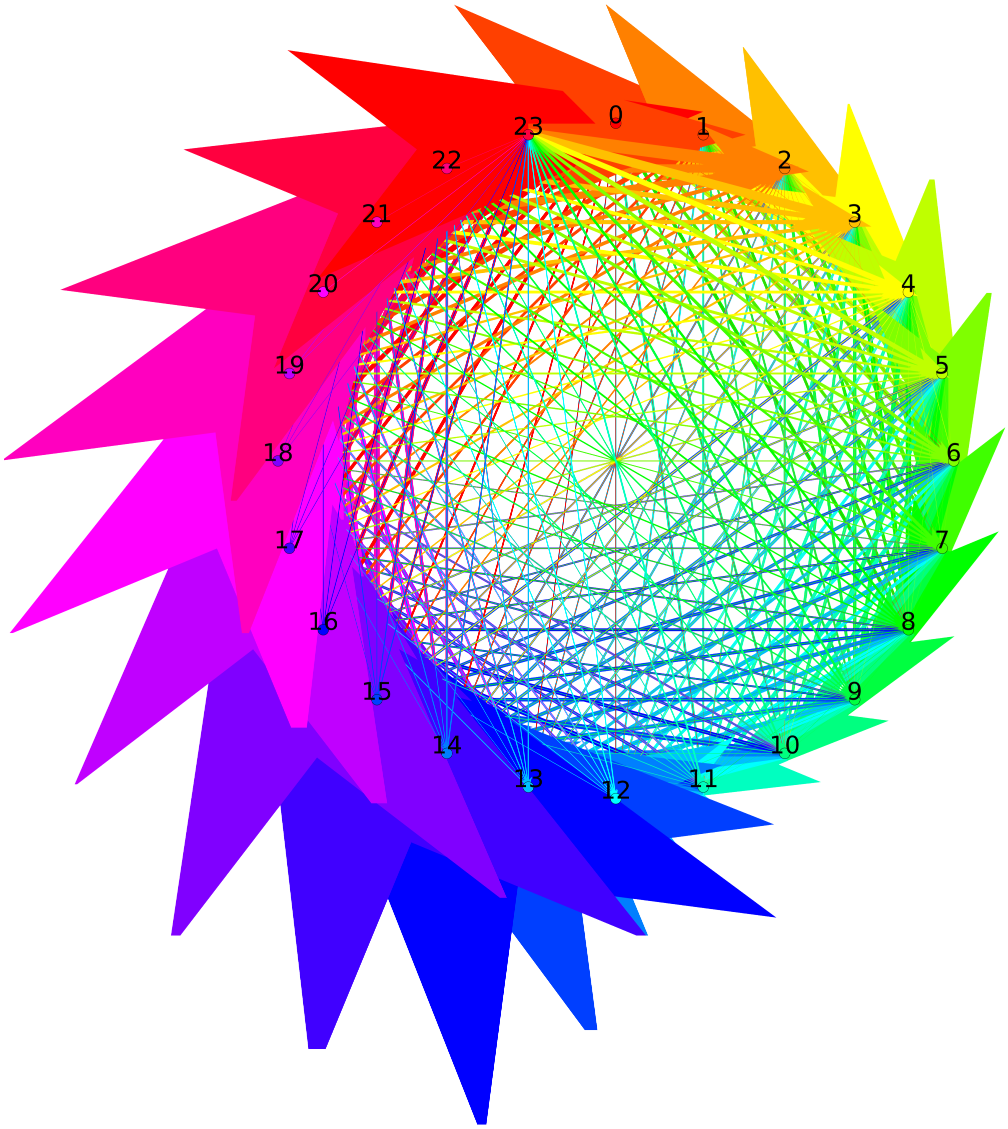

Figure 12: *24-hour posts-clock, showing the time at which posts were made, their volume (as edge- and arrow-size), and the direction of replies (arrows from prompt to reply).*

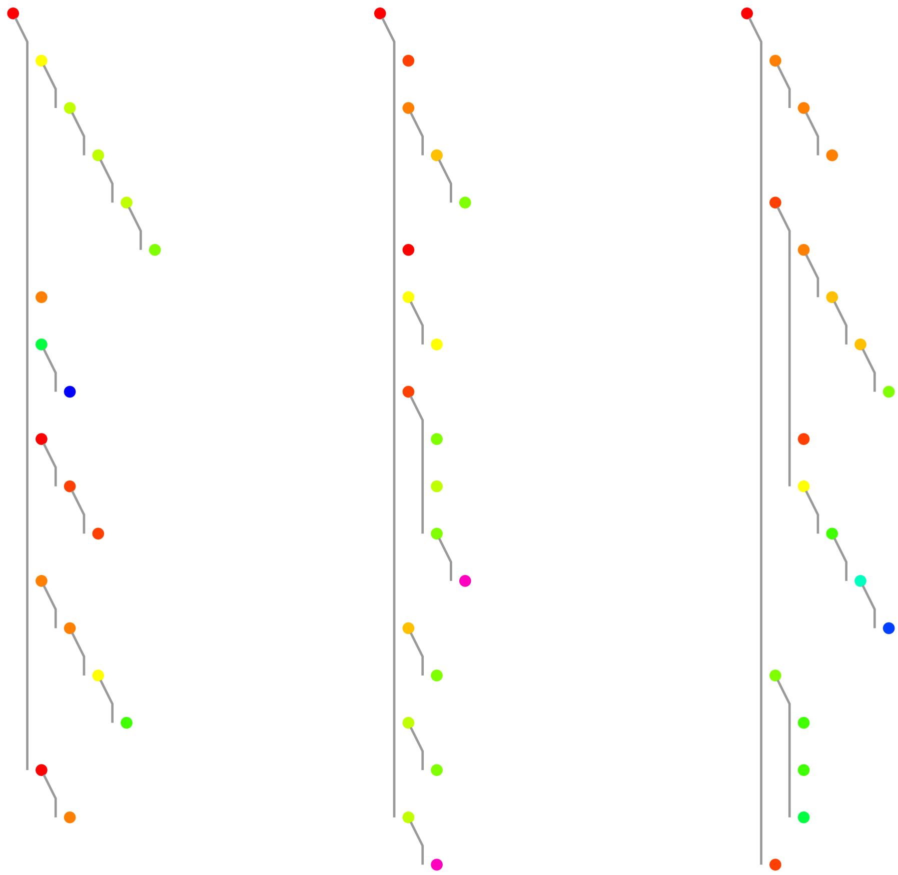

Figure 13: *Progress of time through three randomly selected threads (20-25 posts). Colouring in hours from thread-creation.*

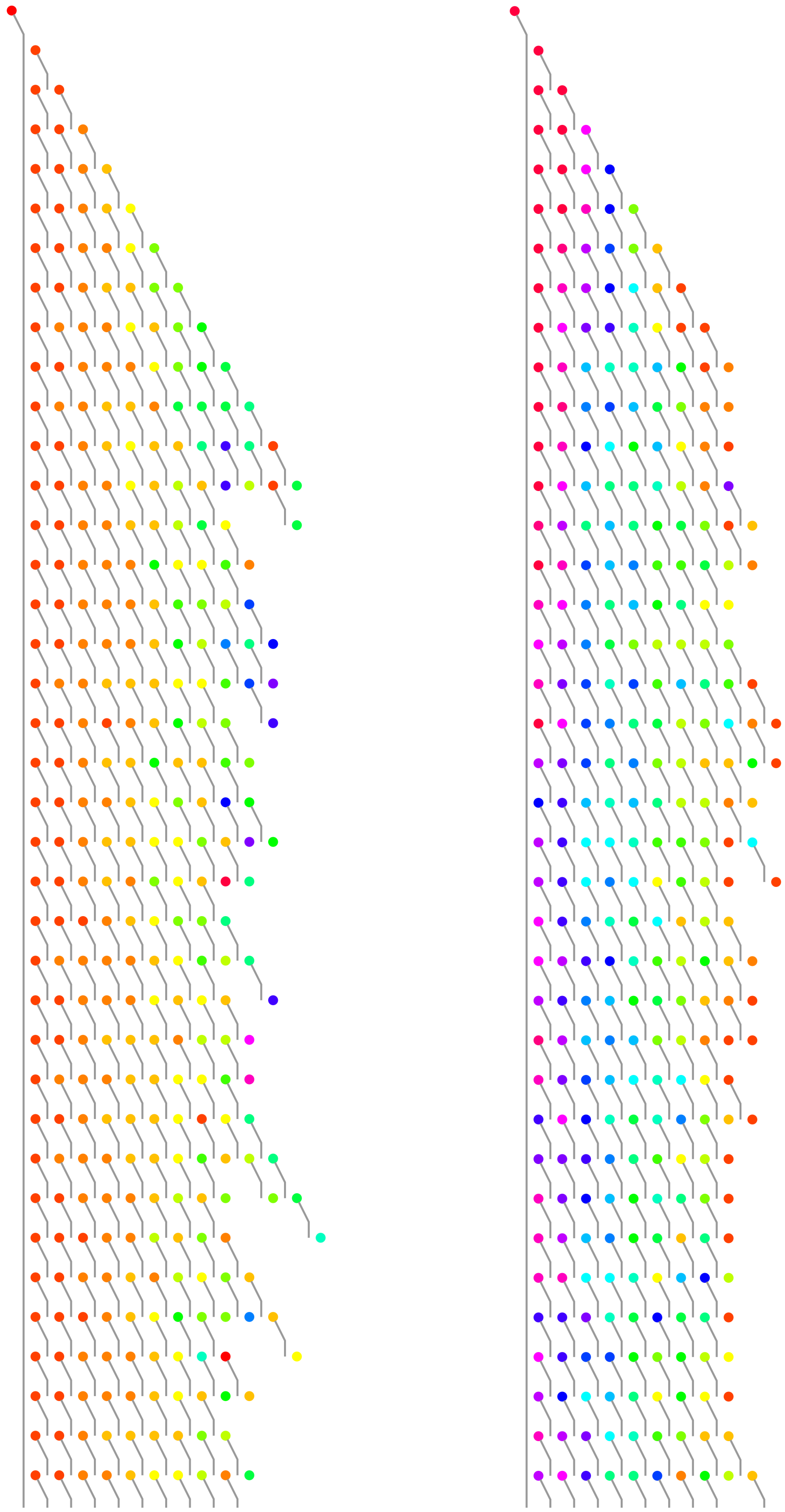

Figure 14: *Heatmaps for the progress of time (left, coloured according to peak windows) and the average karma of posts (right). Sub-threads are (recursively) sorted by time in the latter.*

ordering. As can be observed, earlier posts (top and left) are rated consistently higher, and thus appreciated more.

Though other factors (such as later posts being worse on average) cannot be completely excluded, both the fact that newer posts are rated higher on average, and that posts follow one another in quick succession, seems to indicate that threads are impacted by time-pressures. That is, users mainly reply to newer posts, and therefore the time at which they arrive on HN will determine which posts they can (most productively) reply to. Also, given that the time scale for this social mechanism is measured in hours, circadian effects on social interactions are expected.

3.2 *Representing the social network*

In order to test the second hypothesis, and to analyse time-effects at the level of social relationships, representations of the social network were constructed. A network consists of nodes connected by edges. The nodes represent users. Users are defined as people that post on HN (at least once during the period of investigation). Readers were not considered users here, because they don't actively form part of the community (though they might imagine themselves to be), and cannot form relationships through posting on the site.

Instead of having only users as nodes, there could have been two types of nodes, one representing users and the other threads, resulting in a bimodal network.[33] This was not done for three reasons. First of all, as was shown in the previous section, time-effects operate within threads. Secondly, and foremostly, the impact of time on social relationships is the focus of this study, and reply-relationships are a better proxy for this. And thirdly, threads are shown on the frontpage for quite long, upto a day or more (see figure 16) so any thread-level circadian effects (not captured by reply-relationships between users) should be limited.

This brings us to edges, the relationships between nodes. As just noted, edges represent reply-relationships. That is, if user A replies to user B, they establish a unidirectional reply-tie. For most of the analysis only reciprocated ties were counted, so a reply from B to A would also be required for a bidirectional tie. This requirement was made because social relationships normally imply reciprocity. These bidirectional ties were then encoded as undirected edges in the network. However, as reply-ties also have a direction in time, directed networks maintaining the distinction between ties in each direction were also analysed.

In addition to being (un)directed, edges were also weighted by the number of replies (in each direction). Thus if user A replied to user B four times, the edge was weighted as 4 from A to B. Weighting made it possible to reduce the network in a meaningful way, namely by cutting out edges that represented less than X replies. Where it should be noted that bidirectional relationships were defined as X ($X > 0$) in each direction. The networks will now be discussed.

3.3 *Time in the network*

In the first network under discussion, unidirectional ties of 1 reply were enough for an edge to form (figure 17). It represents the whole reply-structure in our data, and is quite large, with 13,000 nodes and 80,000 edges. It was layed out using OpenOrd, a layout algorithm tailored towards detecting clusters.[32] As can be seen, apart from the main cluster, several smaller ones are visible at the bottom-left. These clusters are mostly uniformly coloured, and this colouring is not provided by the algorithm. Both users and edges were coloured based on time-windows.

Users were coloured based on in which sliding three hour window they normally post. That is, users who do most of their posting between 8:00 and 11:00, are all coloured the same, as are those for 9:00 - 12:00, and so on. Where 'window with most posts' was de-

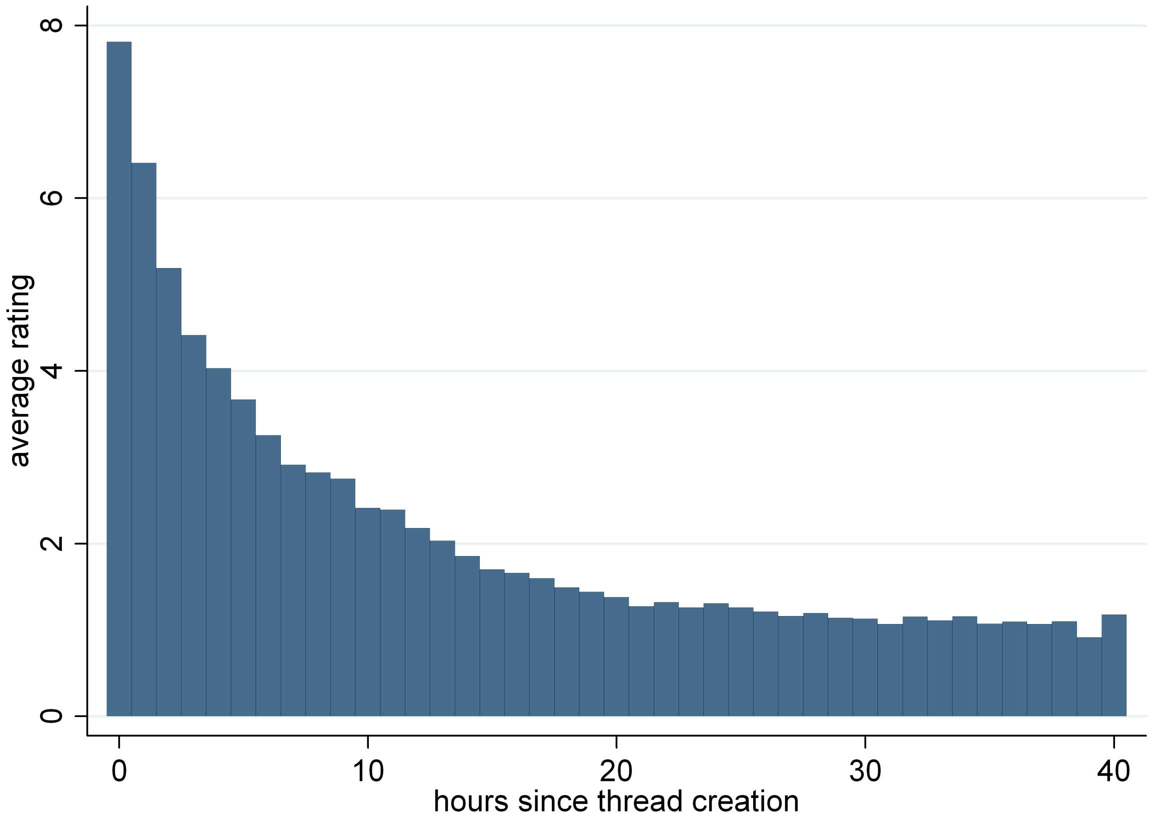

Figure 15: *Average ratings for posts created each hour after thread creation.*

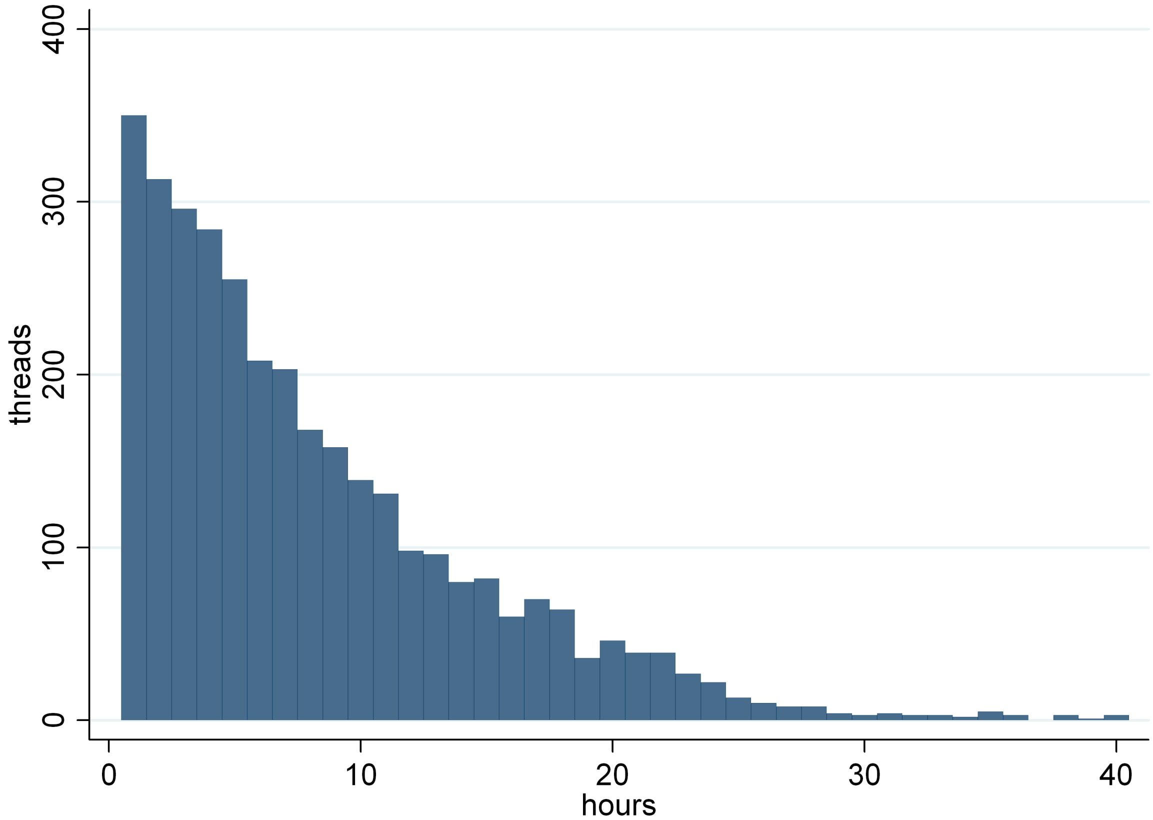

Figure 16: *Time for which threads remain on the frontpage of HN. Total number shown per hour.*

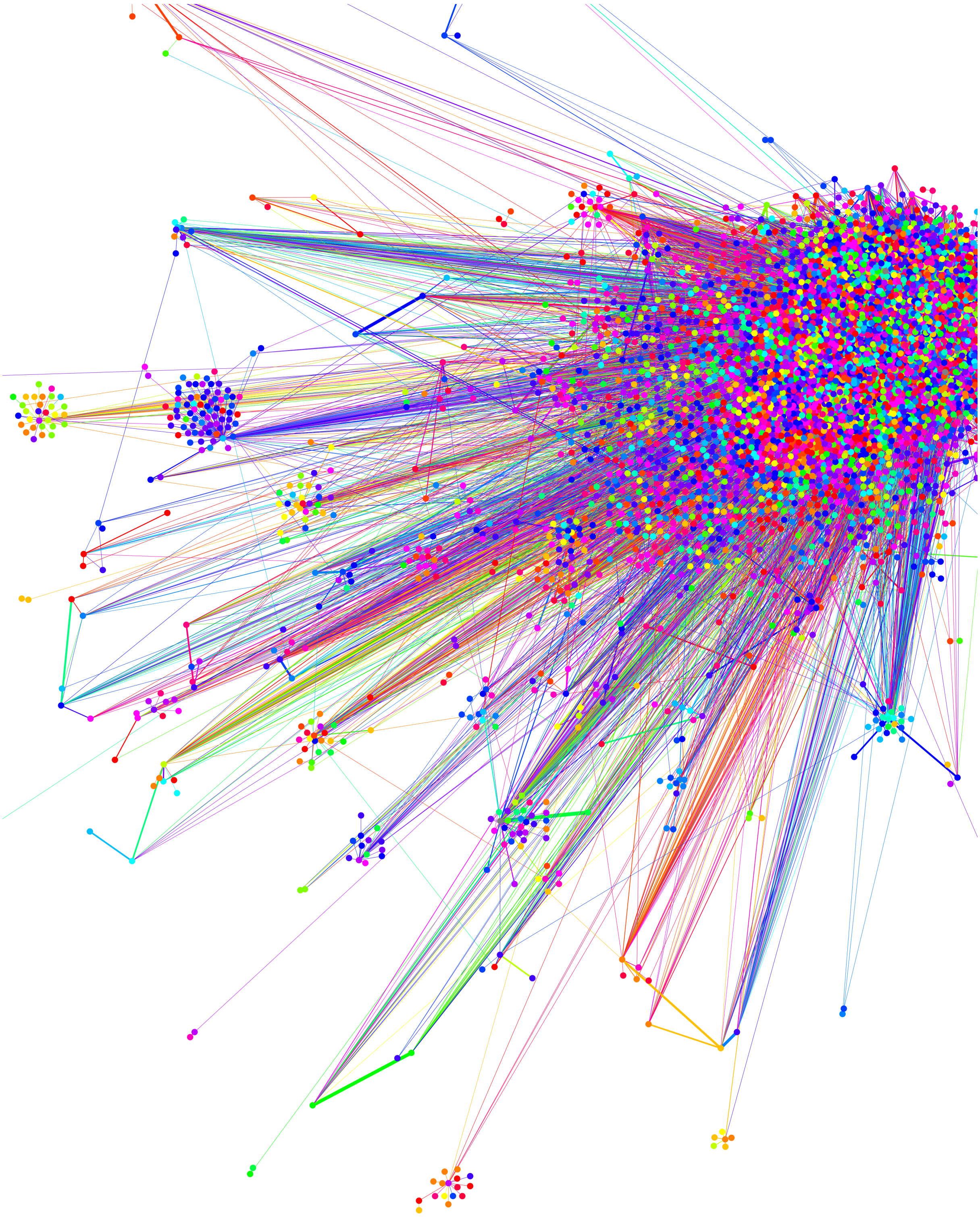

Figure 17: *Complete HN network (for period of investigation) layed out using OpenOrd. Users and edges coloured by window. Tiny time-based clusters can be discerned.*

fined as that with the highest number of posts, and containing at least 1/4th of the users total posts (users without window are shown in grey). Edges were similarly coloured, based on the window in which most of the replies creating that tie were made. The colour-scheme is the same as that used for the posts-clock (figure 12).

As can be seen, the main cluster is a mix of all colours. Breaking it down further was tried using several clustering algorithms, but no large sub-clusters were found. All that could be detected were several dozen small hub-and-spoke 'clusters', of the kind visible here. In these, the spokes are formed by replies from users with only one or a few posts. Most of these are replying to the news-story prompting the thread (level 0 in the thread), rather than replying to replies (level 1 and up). The small clusters mostly disappear if the network is generated with only the latter included.

Weeding the network down by removing all edges representing fewer than 3 replies in each direction, leads to the undirected network shown in figure 18. Rather than it consisting of clusters, one can see users tied together like beads on a string, forming a circle in the middle. Users and edges are here again coloured in the same way. As can be seen, the ties show recurring colour-patterns, especially in the strings that run away from the centre. This is indicative of time-effects, and it seems as if users reply to those that post just before them, receiving replies in turn of those that come online (just) after them on the 24-hour scale.

To get a better view of this, a directed representation of the network was created, maintaining the difference between replies in each direction (figure 19). As can be seen the legs of the bidirectional relationship are sometimes created at different times, indicating recurring differences in the times at these users respond to each others posts (possibly suggesting social ties). However in most cases they are quite close, only having an hour or so between them. Almost the same network, now also including uni-directional links, is shown in figure 20. It has arrows indicating the temporal direction of replies. Though there are some short sequential paths, connecting two or three nodes, no long one-way-paths are visible.

To investigate this further, an users-clock for this network is shown in figure 21. The node in the middle represents users without window. As can be seen, most interaction happens on the left side, and a sizeable amount of it involves users without window. As for the arrows, much of the interaction goes against time. This can in part be explained by windows giving poor coverage of very prolific users (who post during much of the day, and thus can appear to reply to users that come online in the future). But it also suggests only limited time-effects near the centre of the network.

Another user-clock, now for users with five or fewer posts, is shown in figure 22. Windowless users are much less prominent among these, and replies do flow in the right direction here. Figure 23 shows nodes sized by betweenness centrality (the number of paths between nodes going through them). It confirms that the most central users are indeed those without pronounced windows. Interestingly, even for them, edges between specific pairs do display time-patterns.

Thus visual inspection of the networks does seem to uncover some time-effects at the micro-scale. And an even better view of these can be gotten by looking at the video that was made for this paper, of how the directed network shown in figure 19 develops over time (see reference).[52] However, most time-effects seem to involve users at the periphery of the network, and quantitative methods are more suitable for analysing these.

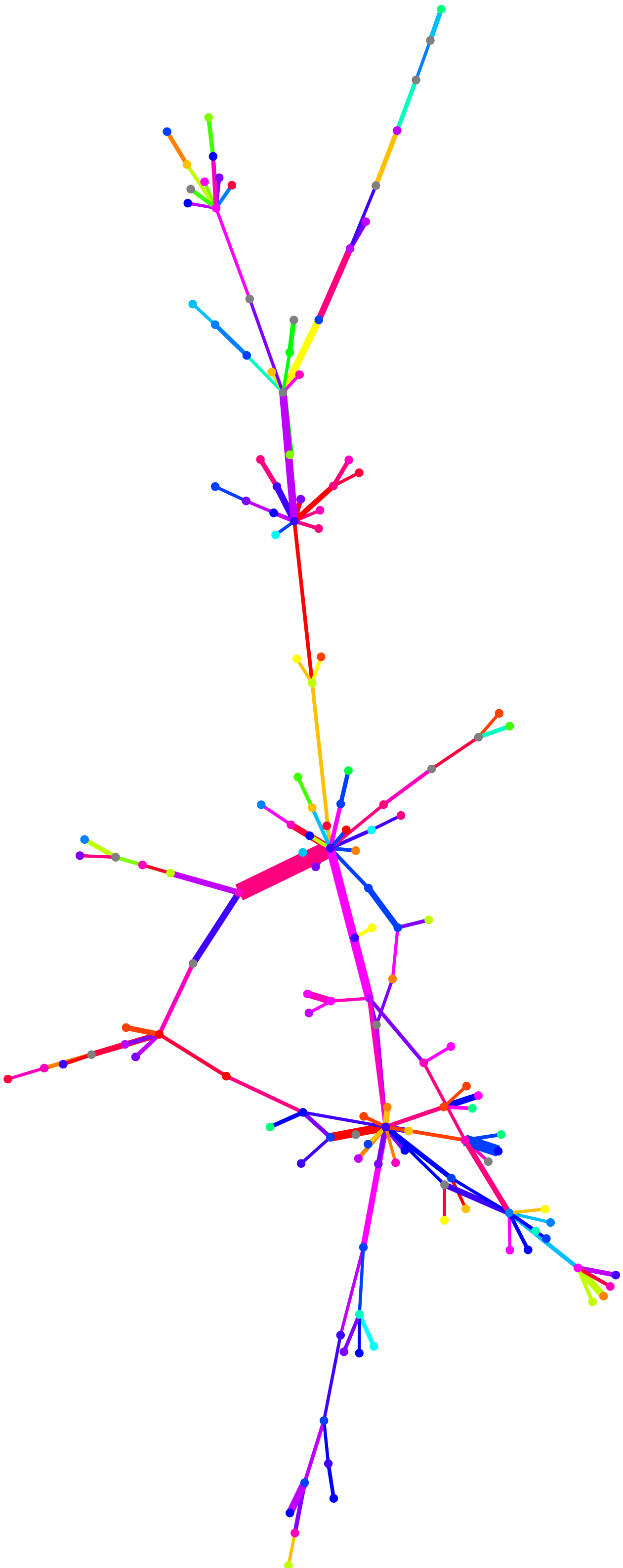

Figure 18: *Undirected, minimum-3-reciprocated-replies network. Users connected by strings of edges (both window-coloured). Recurring time-patterns are visible.*

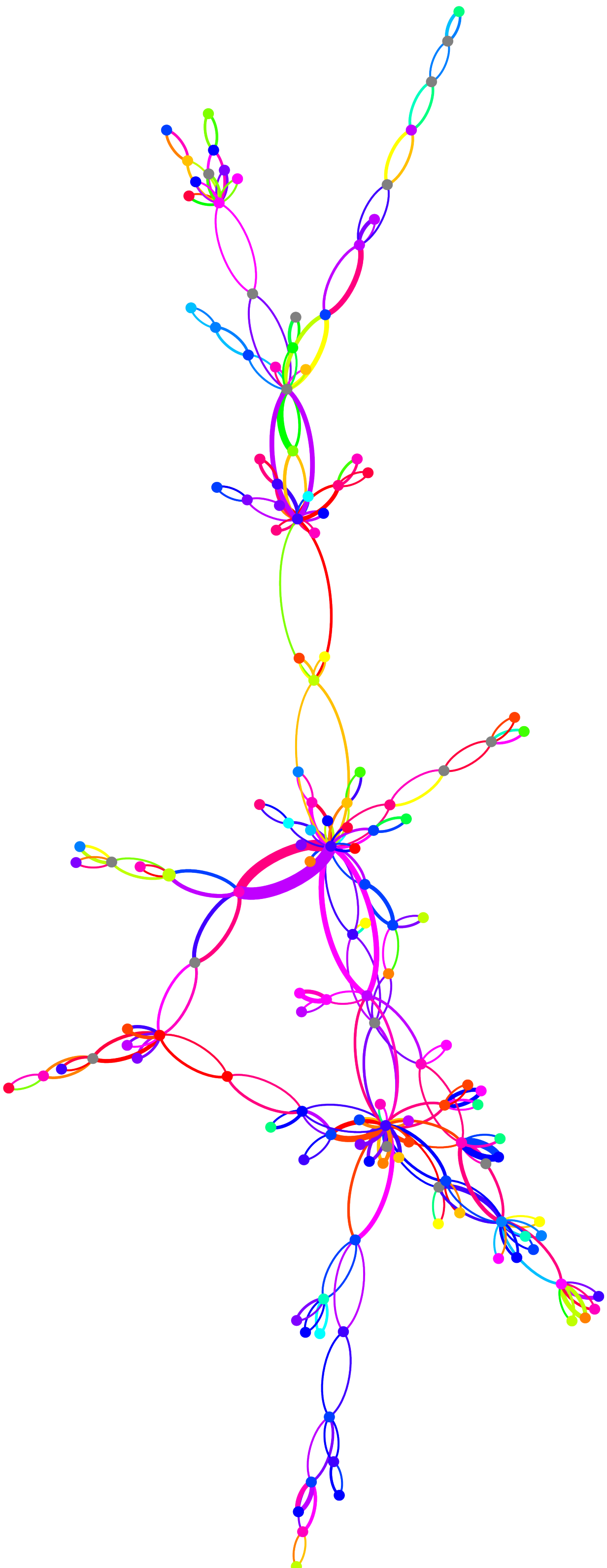

Figure 19: *Directed representation of the network. Time-differences between legs of some relationships visible.*

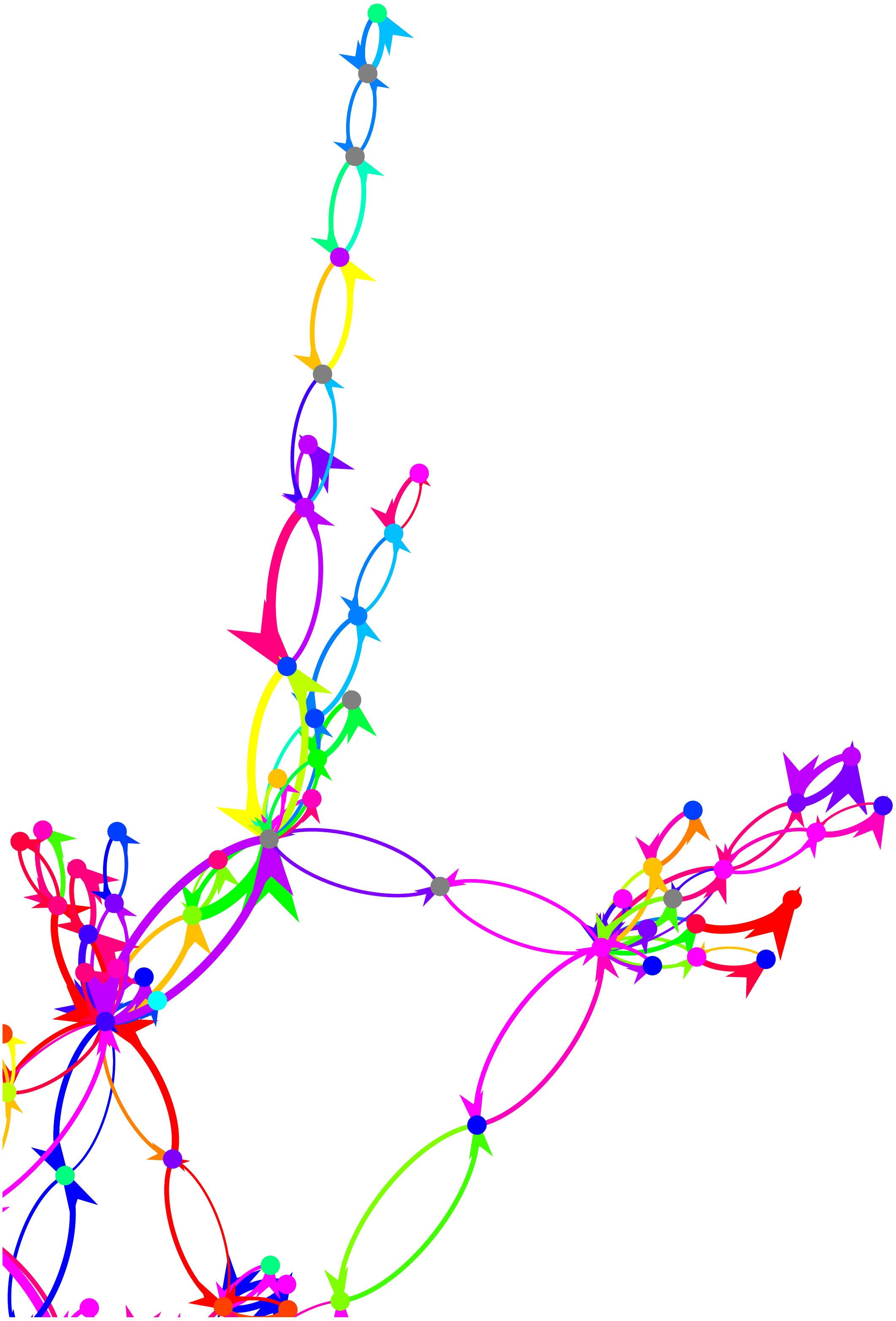

Figure 20: *Directed network with arrows (6 replies minimum, no reciprocity requirement). No long-distance time effects.*

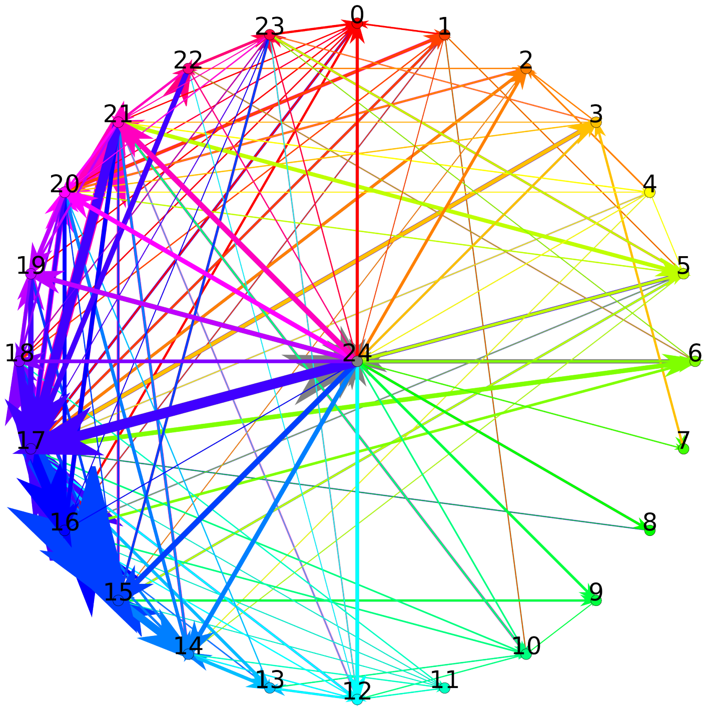

Figure 21: *Users-clock, showing users from the 6-replies network at their peak-window, and replies exchanged by them between peak-windows (thickness represents number of replies). Some arrows point back in time.*

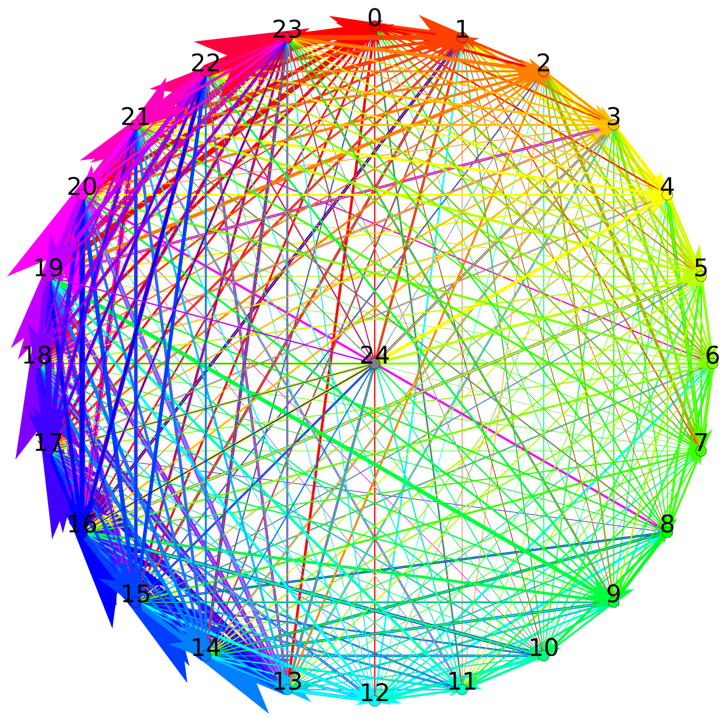

Figure 22: *Users-clock for unprolific users (< 6 posts). Clearer time-effects for these peripheral users.*

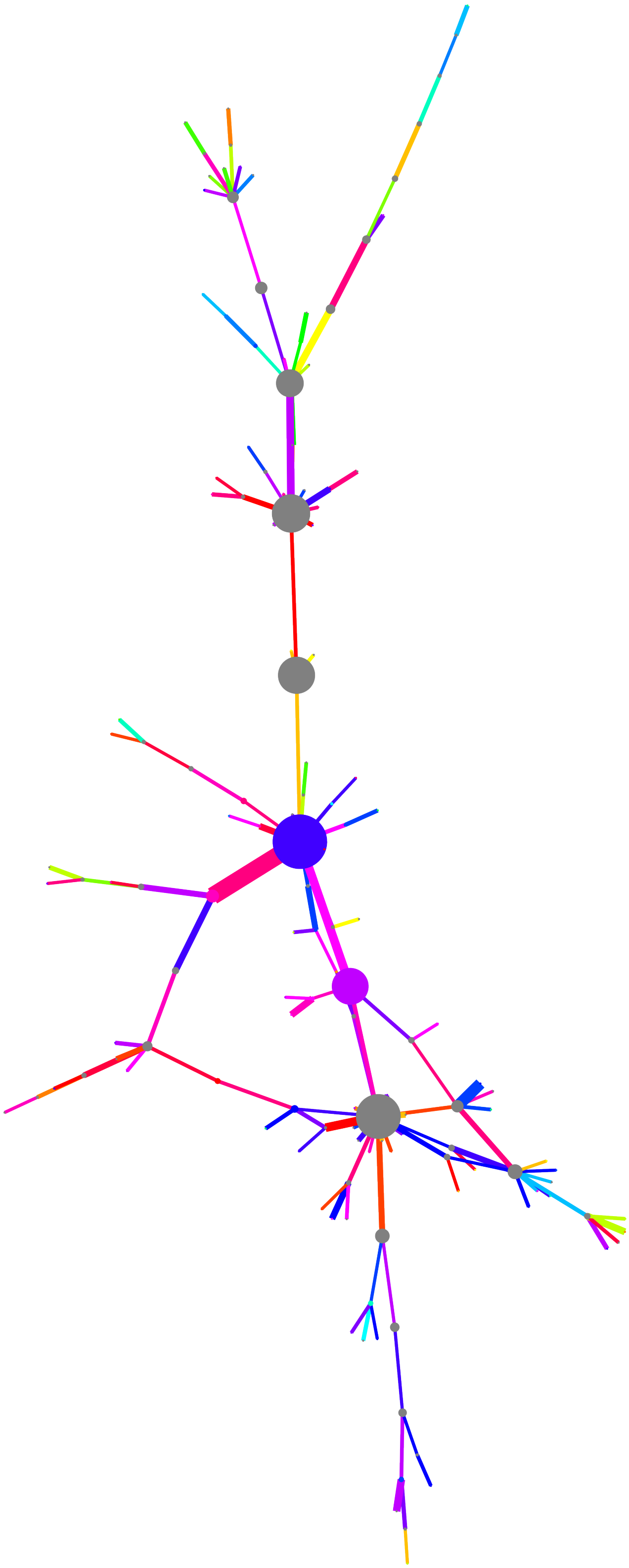

Figure 23: *Users in the original 3-reciprocity network, sized by betweenness-centrality. Windows of 1/3rd rather than 1/4rd of all posts to show unpronounced windows. Large nodes are all grey (or purple/blue), meaning not confined to a window.*

3.4 Measuring time effects

The effect of time was measured in three ways. First of all, relationships formed by replies made within time-windows were compared to those formed in random windows. This comparison was done using reciprocity and transitivity coefficients. Reciprocity measures the fraction of ties that are reciprocated (in a bidirectional network). While transitivity (also known as the global clustering coefficient) is calculated by counting the fraction of triplets (string of three edges) that form triangles. High transitivity indicates triadic closure (friends of friends becoming friends), which is an important proxy for the presence of social relationships.

Making this comparison to random windows, rather than to the values for the whole network (which were higher), was necessary for two reasons: Most importantly, cutting up the network in windows severs ties which otherwise would have contributed to the coefficients (mostly by raising, though possibly lowering them). Secondly, it has been shown that network-size impacts transitivity scores, with smaller networks normally having higher scores. Though this is only true if the smaller network has natural boundaries (is cohesive), and not a random sample drawn from a larger one (then it is lower, because of the randomly severed ties). To exclude all these effects, random time-windows were sampled (rather than the whole network, or random edges). Random time-windows here consisted of a different randomly selected time-window for each of the 40 days, held constant for the sample.

A permutation-test was done using these, creating 10,000 networks for random-window samples, and counting the number of times either measure for such a network was higher than (or the same as) the original value.[56] Where as the original values the lowest transitivity and reciprocity values found for any of the circadian windows were chosen (so very conservative, excluding any remaining effects from size differences). Even then, for random networks transitivity was only higher 41 times, and reciprocity never, leading to p-values of 0.0041, and 0.0000 respectively. From this, with a very high degree of certainty, it can be concluded that both reciprocity and transitivity were higher within circadian time-windows (see table 1).

Table 1: *Minimum & maximum reciprocity and transitivity scores for time-windows, for random windows, and for the whole network. Scores for time-windows were higher than for random windows in a permutation-test, p-values 0.0000 and 0.0041*

Network	Reciprocity	Transitivity
Window min.	0.113	0.016
Window max.	0.148	0.025
Random win.	0.087	0.014
Whole netw.	0.156	0.032

Secondly, a regression was ran over the relationship between network-distance (number of edges passed) and the median circadian time-difference between users posts. The latter would be 5 hours if user A posts at 23:00, and user B at 4:00. The results of this regression were quite limited, as can be seen in table 2. A time-difference of only ten minutes per hop (between 1 and 4 hops, it levels off after that) was found for the core of the network (users with 3 reciprocated ties minimum), while it was a bit over half an hour in the periphery (users with 5 or less posts). Especially the R-square values are quite low (only 1 to 2% of variance explained). However, given that no clear clusters were found, a weak relationship between time and network-distances is not surprising.

Thirdly, and finally, the effect of the phased introduction of daylight saving time (DST) on the network was measured. A comparison was made between the two weeks before DST (in the US and UK), and after the US went on DST (13th of March) but the UK was

Table 2: *Regressions of network-distance to time. Values are time in minutes per hop (over hops 1-4). Results are highly significant, but Rsquare values are very low (only 1-2% of variance explained).*

Network	Coef.	StdE.	P-val.	95% CoInf.	Rsq.
3-Recip.	13.66	2.23	0.000	9.29 – 18.04	0.020
Unprolf.	32.22	6.10	0.000	20.25 – 44.18	0.018

Table 3: *Average distance between UK and US communities before and after US DST. P-value for t-test would be 0.0000*

DST	Mean Dist.	StdE.	95% CoInf.
Before	3.65	0.017	3.62 – 3.69
After	3.53	0.015	3.50 – 3.56
Diff.	-0.12	0.019	-0.08 – -0.16

not yet (until the 27th). Networks for both 2-week periods were generated, and within these two networks, network-distances between US-Westcoast (Los Angeles, San Francisco) and UK users were calculated. As expected, the distance was smaller after the introduction of US-DST (8-hour difference temporarily reduced to 7 hours) (see table 3).

Finding this reduction in distance is not only indicative of time-effects, but it also serves as evidence against the hypothetical sceptic who would claim that all time-effects are due to spatial distances: as DST is purely a shift in time. If the pre and post-DST-distances were to be ran through a paired t-test, they would be found to be very significant ($p < 0.0000$). However, a t-test can not be reliably applied here, as the observations (nodes) are interdependent even under the null-hypothesis. This because in the worst case only a few edges could have reduced the distance between many nodes. Given the fact that not many clusters were found, the odds of this are small, but that does not change the fact that more advanced statistical methods are called for here (that take this chance into account).

Also, for all these measures (as well as for the visualizations) it should be noted again that no controls were added for possible confounding factors, such as geographical distance, or users age and lifestyle. This was not deemed a priority for a first study into time-effects in forums, especially as this information was not readily available for all/sufficient data-points. But on to the discussion now.

4 Discussion

Coming back to the two hypotheses, the first one, concerning the existence of time-pressure effects in threads, can be tentatively confirmed. That is, most replies are prompted by messages less than two hours of age, leading to a daily reply-circle, as represented in the 24-hour posts clock. Time also plays an important role in the overall structure of the thread, as visualized in the heatmaps. Timeliness is thus indeed important within threads, and at the level of individual posts.

The second hypothesis, that the time-pressure effect would shape the social structure of the community, was not fully substantiated, though some evidence was found.

In the core, time-effects were small. Micro-level effects were visualized, such as the string-like chains of users from successive time-windows, and different legs of the relationships being created at different times. But though real (especially for edges), these did not translate into much quantitatively, as in the core at best every hop across the network accounted for a circadian difference of 10 minutes.

As for why time-effects were not found to be strong in the core, two main explanations can be brought forward: It might be true that most of the socializing happens between active users at the core, where they get to know one another, and actively reply to their 'friends', even if they posted hours ago. An alternative explanation could be that as active users, they visit the site more often, and write more posts (they have wider peaks, if at all). Which then makes them more likely to form reciprocated ties with other (especially active) users, and moves them to the core of the network. Which of these explanations is closer to the truth, is hard to say at this stage.

In the periphery, however, time-effects were slightly more pronounced. The time per hop was 30 minutes there (though with much variance unaccounted for), and the small time-based clusters that could be found, all involved peripheral users. Also, on a macro-level, reciprocity and transitivity were proven to be higher within time-windows than in the rest of the network. Finally, the phased introduction of DST did have a significant effect, but it was quite small again. Thus to the extent that time-pressure effects shape the social community, they do so very gently, and at the edges.

Linking it all back to the theory, the social mechanism of replying under time-pressure was found to be at work in threads, and even quite strongly so. However this did not translate into very large effects at the aggregate, emergent level. These findings do not provide (strong) support for Structuration Theory, nor for the social mechanisms-based approach. Though they cannot be held against them either. Most likely other mechanisms interfered, such as for example the fact that active users are less confined to time-windows, and interact with a lot of users, making it likely that they bridge the gaps that time creates. Alternatively, it might be true that (time-based, or social) clusters only form when users receive sufficient social cues (such as avatar images), and that HN, with its relatively sober layout, did not provide enough of these.[37,2]

5 Conclusion

To conclude, two hypotheses were formulated based on theory: 1) the existence of a time-pressure mechanism favouring timely replies at the thread-level, and 2) this mechanism impacting the social structure of the HN community. The first was confirmed, while the second was not fully. The main finding being that to the extent time-effects impact the social structure of HN, they are weak in the core, where they only can be observed at the micro-level, while in the periphery, and on an aggregate level, they are slightly stronger.

In other words, the hands of time might set the stage, and be behind some of the choreography, but they certainly do not direct the play. Of course, as a tentative mapping of the issue, this paper did not do more than give a glimpse of the formerly invisible hands of time. Catching time red-handed, or even being able to exclude all alternative explanations, will require more work, more advanced statistical methods (some of which have not been developed yet), and most likely, an on-line experiment.

Finally, as this was only a study into the effect of time on the social structure of one specific thread-based online community, it would be interesting to cross-validate these findings using different data-sets, such as different news-sites, forums, Facebook wall-comments, or even journal papers across differently structured academic years. Even if small and subtle, time-effects might be at work in many other settings, suggesting unexplored directions, bringing together co-founders with similar waking patterns, and moulding our social relationships, right now, and every day.

6 Acknowledgements

I would like to express my thanks to Dr. Grant Blank for his advice on statistical methods.

Bibliography

1. Anderson, B. R., *Imagined communities: Reflections on the origin and spread of nationalism.* (Verso Books, 2006).
2. Chen, G., & Chiu, M. M, 'Online discussion processes: Effects of earlier messages' evaluations, knowledge content,

social cues and personal information on later messages.', *Computers & Education* 50, 678–692 (2008).

3. Chen, H., Shen, H., Xiong, J., Tan, S., & Cheng, X., 'Social network structure behind the mailing lists.', *Proceeddings of the 15th Text REtrieval Conference (TREC 2006)* (2006).
4. Choi, J. H, & Danowski, J. A, 'Making a global community on the net–global village or global metropolis?: A network analysis of usenet newsgroups.', *Journal of Computer-Mediated Communication* 7 (2002).
5. Crandall, D., Cosley, D., Huttenlocher, D., Kleinberg, J., & Suri, S., 'Feedback effects between similarity and social influence in online communities.', *Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining* 160–168 (2008).
6. Curtis, P., 'Mudding: Social phenomena in text-based virtual realities.', *High noon on the electronic frontier: Conceptual issues in cyberspace* 347–356 (1996).
7. Dagon, D., Zou, C., & Lee, W., 'Modeling botnet propagation using time zones.', *Proceedings of the 13th Annual Network and Distributed System Security Symposium (NDSS'06)* (2006).
8. Fisher, D., Smith, M., & Welser, H. T, 'You are who you talk to: Detecting roles in usenet newsgroups.', *System Sciences, 2006. HICSS'06. Proceedings of the 39th Annual Hawaii International Conference on* 3, 59b (2006).
9. Garton, L., Haythornthwaite, C., & Wellman, B., 'Studying online social networks.', *Journal of Computer-Mediated Communication* 3 (1997).
10. *Gephi, an open source graph visualization and manipulation software.* <http://gephi.org/>.
11. Giddens, A., *The constitution of society: Outline of the theory of structuration.* (University of California press, 1984).
12. Gómez, V., Kaltenbrunner, A., & López, V., 'Statistical analysis of the social network and discussion threads in slashdot.', *Proceeding of the 17th international conference on World Wide Web* 645–654 (2008).
13. Graham, P., *Hacker News news.* <http://ycombinator.com/newsnews.html>.
14. *Hacker News.* <http://news.ycombinator.com/>.
15. *Hacker News | How Hacker News ranking algorithm works.* <http://news.ycombinator.com/item?id=1781013>.
16. Hansen, D., Shneiderman, B., & Smith, M. A., *Analyzing social media networks with NodeXL: insights from a connected world.* 1st ed. (Morgan Kaufmann, 2010).
17. Hedstrom, P., 'Explaining the growth patterns of social movements.', *Understanding Choice, Explaining Behaviour* (2006).
18. Hedstrom, P., & MyiLibrary, *Dissecting the social: on the principles of analytical sociology.* (Cambridge University Press Cambridge, UK, 2005).
19. Hedstrom, P., & Ylikoski, P., 'Causal mechanisms in the social sciences.', *Annual Review of Sociology* 36, 49–67 (2010).
20. Himelboim, I., 'Civil society and online political discourse: The network structure of unrestricted discussions.', *Communication Research* (2010).
21. Kaltenbrunner, A., Gomez, V., & Lopez, V., 'Description and prediction of slashdot activity.', *Web Conference, 2007. LA-WEB 2007. Latin American* 57–66 (2007).
22. Kaltenbrunner, A., Gonzalez-Bailon, S., & Banchs, R., 'Communities on the web: Mechanisms underlying the emergence of online discussion networks.', (2009).

23. Kaltenbrunner, A. *et al.*, 'Homogeneous temporal activity patterns in a large online communication space.', *IADIS International Journal on WWW/INTERNET* 6, 61–76 (2008).
24. Karnstedt, M., Hennessy, T., Chan, J., & Hayes, C., 'Churn in social networks: A discussion boards case study.', (2010).
25. Kelly, J. W, Fisher, D., & Smith, M., 'Friends, foes, and fringe: norms and structure in political discussion networks.', *Proceedings of the 2006 international conference on Digital government research, May* 21–24 (2006).
26. Kraut, R. E., & Resnick, P., *Evidence-based social design: mining the social sciences to build online communities.* (2011).
27. Lampe, C., & Johnston, E., 'Follow the (slash) dot: Effects of feedback on new members in an online community.', *Proceedings of the 2005 international ACM SIGGROUP conference on Supporting group work* 11–20 (2005).
28. Li, W. J, Hershkop, S., & Stolfo, S. J, 'Email archive analysis through graphical visualization.', *Proceedings of the 2004 ACM workshop on Visualization and data mining for computer security* 128–132 (2004).
29. Lieblein, E., 'Critical factors for successful delivery of online programs.', *The Internet and Higher Education* 3, 161–174 (2000).
30. Maier, G., Feldmann, A., Paxson, V., & Allman, M., 'On dominant characteristics of residential broadband internet traffic.', *Proceedings of the 9th ACM SIGCOMM conference on Internet measurement conference* 90–102 (2009).
31. Marbach, J., *Hacker News map.* <http://jmarbach.com/hackernews>.
32. Martin, S., Brown, W. M, Klavans, R., & Boyack, K. W, 'OpenOrd: An open-source toolbox for large graph layout.', *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series* 7868, 4 (2011).
33. Newman, M., *Networks: An introduction.* (Oxford Univ Pr, 2010).
34. *NodeXL: Network overview, discovery and exploration for Excel.* <http://nodexl.codeplex.com/>.
35. *Nokogiri.* <http://nokogiri.org/>.
36. *Ohloh, the open source network.* <http://www.ohloh.net/>.
37. Opsahl, T., & Hogan, B., 'Growth mechanisms in continuously-observed networks: Communication in a Facebook-like community.', *Forthcoming* (2011).
38. Orlikowski, W. J, 'The duality of technology: Rethinking the concept of technology in organizations.', *Organization science* 3, 398–427 (1992).
39. Osborne, L. N, 'Topic development in usenet newsgroups.', *Journal of the American Society for Information Science* 49, 1010–1016 (1998).
40. *Pajek.* <http://vlado.fmf.uni-lj.si/pub/networks/pajek/>.
41. Paolillo, J., 'The virtual speech community: social network and language variation on IRC.', *Journal of Computer-Mediated Communication* 4, 0–0 (2006).
42. Ren, Y., Kraut, R., & Kiesler, S., 'Applying common identity and bond theory to design of online communities.', *Organization Studies* 28, 377 (2007).
43. *Slashdot: News for nerds, stuff that matters.* <http://slashdot.org/>.
44. Smith, M. A, 'Netscan: Measuring and mapping the social structure of usenet.', (1997).
45. *Stata: Data analysis and statistical software.* <http://www.stata.com/>.
46. Stein, E., & Daude, C., 'Longitude matters: Time zones and the location of foreign direct investment.', *Journal of International Economics* 71, 96–112 (2007).

47. Surmacz, T. R, 'Measurement of data flow in usenet news management.', *Measurement Science Review* 47–51 (2005).
48. *The igraph library for complex network research.* <http://igraph.sourceforge.net/>.
49. Tsagkias, M., Weerkamp, W., & Rijke, M. D., 'Predicting the volume of comments on online news stories.', *Proceeding of the 18th ACM conference on Information and knowledge management* 1765–1768 (2009).
50. Turner, T. C, & Fisher, K. E, 'The impact of social types within information communities: Findings from technical newsgroups.', (2006).
51. Turner, T. C, Smith, M. A, Fisher, D., & Welser, H. T, 'Picturing usenet: Mapping computer-mediated collective action.', *Journal of Computer-Mediated Communication* 10, 7 (2005).
52. *Visible hands of time video - The social network of Hacker News.* <http://wybowiersma.net/pub/hn_network.avi>.
53. Wellman, B., 'An electronic group is virtually a social network.', *Culture of the Internet* 179–205 (1997).
54. Wellman, B., 'Computer networks as social networks.', *Science* 293, 2031 (2001).
55. Welser, H. T, Gleave, E., Fisher, D., & Smith, M., 'Visualizing the signatures of social roles in online discussion groups.', *Journal of Social Structure* 8 (2007).
56. Wiersma, W., Nerbonne, J., & Lauttamus, T., 'Automatically extracting typical syntactic differences from corpora.', *Literary and linguistic computing* 26, 107 (2011).
57. *wybo/Forum-Tools - GitHub.* <https://github.com/wybo/Forum-Tools>.
58. Yamakami, T., 'A mobile clickstream time zone analysis: Implications for realtime mobile collaboration.', *Knowledge-Based Intelligent Information and Engineering Systems* 855–861 (2004).
59. Zhongbao, K., & Changshui, Z., 'Reply networks on a bulletin board system.', *Physical Review E* 67, 36117 (2003).