

Community structure and interaction dynamics through the lens of quotes

Mattia Samory
Università degli Studi di
Padova
samoryma@dei.unipd.it

Federica Bogo
MPI Intelligent Systems/
Microsoft Research
febogo@microsoft.com

Enoch Peserico
Università degli Studi di
Padova
enoch@dei.unipd.it

ABSTRACT

This is the first work investigating community structure and interaction dynamics through the lens of quotes in online discussion forums. We examine four forums of different size, language, and topic. Quote usage, which is surprisingly consistent over time and users, appears to have an important role in aiding intra-thread navigation, and uncovers a hidden “social” structure in communities otherwise lacking all trap-pings (from friends and followers to reputations) of today’s social networks.

CCS Concepts: •Human-centered computing → Social networking sites;

Keywords: quotes, implicit social network, forum, thread

Acknowledgement: supported by MIUR Proj. AMANDA

1. INTRODUCTION

We examine four online forums of different size, language and topic through the lens of *quotes* – excerpts from previous posts that a new post can cite. We show that quotes have a role in navigating discussion, and reveal the social structure of the debating community. Quotes are a widespread mechanism in forums, yet one mostly disregarded by academic literature. This paper extends our research [4], generalizing our findings to different forums, and quantifying the preliminary intuition that quotes help sustain long discussions and reveal the latent structure of the community.

Quotes are precursors of affordances such as shares, cites, and retweets, in the same roles as discussion aggregators and signals of attribution, acknowledgement, and endorsement. In modern social networks these signals can be concurrently expressed through alternative affordances (e.g. likes), and are often subordinate to prior befriending or following. The fact that online forums typically lack or see little use of other “social” mechanisms makes the analysis of quotes less noisy.

We remark that quotes differ from *replies* (analysed e.g. in [1]). Replies link each post beyond the first to *exactly* one previous post *in the same thread*, effectively organizing the

	RPG	SWZ	TM	PSY
<i>posts</i>	14.3M	1M	3.6M	0.15M
<i>users</i>	56.9K	29.9K	14.9K	2.8K
<i>threads</i>	522.7K	112.1K	49.2K	24K
<i>quotes</i>	8.4M	21.8K	1.6M	31.1K
<i>timespan (years)</i>	'00-'13	'02-'14	'01-'14	'04-'14
<i>language</i>	EN	IT	IT	EN

Table 1: Base statistics for the crawled forums.

thread into a tree of posts rather than into a linear sequence. Unlike replies, quotes allow a post to link multiple previous posts (or none), potentially belonging to other threads or even subforums. Furthermore, quotes explicitly identify the portion of the linked post to which they refer.

2. DATA

We crawled a set of four forums, acquiring all posts available since each forum’s inception. The data gathered spans more than 10 years for all forums. The forums differ both quantitatively (see Table 1) and qualitatively – in terms of goal, focus, and typical evolution of discussion. The forums analysed are:

RPG - the largest international online forum devoted to roleplaying games (*RPGs*), with a focus on tabletop *RPGs*. <http://forum.rpg.net>

SWZ - the forum section of an Italian IT news and information website. <http://forum.swzone.it>

TM - a major Italian board for discussing metal and hard rock music. <http://truemetal.it/forum>

PSY - self-defined mental health support community. <http://forum.psychlinks.ca>

3. DISCUSSION THROUGH QUOTES

This section quantitatively characterizes the presence and role of quotes in discussion. Quote usage varies across forums, with the quotes/posts ratio ranging from 0.587 in *RPG* to 0.022 in *SWZ*. In all four forums, however, the ratio appears to follow power-law distributions. Interestingly, in each forum the power law exponent for quotes made *to* a post almost perfectly matches that for quotes *by* a post.

Another characteristic of quotes is that *in each forum the ratio of quotes/posts remains constant over time and users*. This is remarkable, given that not only does the post count change significantly from month to month, but that the average user “lifetime” (less than 1.5 years for all four forums)

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

WebSci '16 May 22-25, 2016, Hannover, Germany

© 2016 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-4208-7/16/05.

DOI: <http://dx.doi.org/10.1145/2908131.2908195>

	RPG	SWZ	TM	PSY
Nodes	35118	11544	9661	1553
Edges	2.5M	50.8K	291.7K	5983
Zero InDeg Nodes	3330	1832	533	117
Zero OutDeg Nodes	9628	6084	2996	853
NonZero Deg Nodes	22.2K	3628	6132	583
Unique directed edges	2.5M	50.8K	291.7K	5983
Unique undirected edges	1.8M	41.5K	203.9K	4804
Self Edges	4174	766	1265	70
BiDir Edges	1.4M	19.3K	176.8K	2.4K
Closed triangles	176.8M	250K	5.4M	11.4K
Open triangles	1.4G	11.9M	51.2M	715K
Frac. of closed triads	0.111	0.021	0.0962	0.0158
Connected comp. size	0.995	0.975	0.993	0.993
Strong conn. comp. size	0.625	0.283	0.625	0.365
Approx. full diameter	7	7	7	6
90% effective diameter	3.317	3.690	3.338	2.922
Average clustering	0.385	0.305	0.469	0.431
Assortative mixing	0.088	-0.249	0.226	-0.004

Table 2: Quote network statistics.

is significantly shorter than the time interval under observation – meaning that each forum’s signature quotes/posts ratio is adopted by generations of posters. It is also notable that more prolific posters receive (and make) more quotes, but no more and no less than groups of less prolific users with the same total post count – *there is no “rich-get-richer” effect*, in marked contrast to most other social environments.

Even though quotes by/to a user precisely track that user’s post count, the same cannot be said of quotes by/to a thread. Short threads both make and receive relatively fewer quotes per post. A possible explanation is that *quotes aid navigating discussions* – with shorter threads being intrinsically easier to navigate and thus requiring less quote support.

Further analysis supports this hypothesis. First, quote length markedly grows with the temporal distance between quoting and quoted post. Most of the shortest quotes (e.g., “no”, “3?”, “me”) refer to a very “close” post on which they rely to provide the appropriate context. Second, *quotes are efficient shortcuts in long discussions*. Define the *depth* of the first post in a thread as 0, and that of any other post p as 1 plus the minimum depth of any post that p quotes or immediately follows in the thread. Quotes decrease the average and maximal post depth of longer threads significantly more than that of short ones, in all four forums; and forums with longer threads (*RPG* and *TM*) exhibit more quotes and greater reduction of thread depth than forums with shorter threads (*SWZ* and *PSY*).

4. COMMUNITY THROUGH QUOTES

Online forums do not show which users, or even how many users, quote a given user or post. It may then be surprising that the latent structure of a forum’s community can be retrieved by observing how users quote one another, and that this structure shows the typical features of a social network.

Consider the *author-quote graphs* of the four forums – the directed, weighted graphs where each node is an author and an edge between nodes a and b is weighted by the number of times a quotes b . Table 2 shows that these graphs sport many characteristics typical of social networks. First, the graphs are sparse, containing only a small fraction of all po-

tential edges. Second, they are small worlds, with a giant connected component. The number of nodes in the weakly connected components of all forums, and strongly connected components of *RPG* and *TM*, closely match the corresponding values for the Twitter network (respectively 92% and 68% [2]); the strongly connected components of *SWZ* and *PSY* are slightly smaller. Furthermore, the diameters for the largest components are relatively small.

Quotes are highly reciprocated: roughly 50% of all node pairs connected by an arc sport an arc in the opposite direction. The clustering coefficient, too, is remarkably high, and remains high even for nodes of high degree, more than in the Twitter or Facebook graphs [2] – a possible explanation lying in the highly specialized nature of forums that tends to limit the variety of a user’s circles. Finally, assortativity by node degree is mildly positive for *RPG* and *TM* (like in the Facebook or Twitter graphs [2]), and mildly negative for *PSY* and *SWZ* (as the Internet and WWW graphs [3]).

The differences between *RPG* and *TM*, and *SWZ* and *PSY*, match the intuition of the first pair of forums being driven by more peer-to-peer conversations, and the second pair being venues for obtaining information from experts.

Preliminary results show evidence that quote networks are indeed social in a more fundamental way. Quote usage seems dependent on the identity of users, rather than just on discourse. To verify this claim, we used parameters of users’ ego networks to identify them across distinct sets of discussions – a simple classifier can distinguish two users with over 80% accuracy in any of the four forums. Moreover, quotes appear to be reliable indicators of bonding between users: leveraging quote network metrics, we can recover friendship relations established through the forums’ rarely used friendship mechanism with $\approx 70\%$ average accuracy. The full version of this work [5] provides all details.

5. CONCLUSIONS

Quote usage, that is surprisingly consistent over time and users in each of the four forums we examined, appears to play an important role in aiding discussion navigation, and reveals the structure of the user community. Preliminary results also suggest that quotes are in some sense weak “digital fingerprints”. Future work should explore to what extent they can be leveraged to identify users, and characterise their behaviours, within a forum – or even across different ones. Quotes are “higher resolution” tools than likes, shares, and cites – in this sense it would be interesting to see if the information they provide can still be recovered in networks that only offer the latter lower resolution tools.

6. REFERENCES

- [1] E. Aumayr, J. Chan, and C. Hayes. Reconstruction of threaded conversations in online discussion forums. In *Proc. ICWSM*, 2011.
- [2] S. Myers, A. Sharma, P. Gupta, and J. Lin. Information network or social network? The structure of the Twitter follow graph. (*Companion*) *Proc. WWW*, 2014.
- [3] M. Newman and J. Park. Why social networks are different from other types of networks. *Phys. Rev. E*, 2003.
- [4] M. Samory and E. Peserico. Quotes in forum.rpg.net. In *Proc. WebSci*, 2015.
- [5] M. Samory and E. Peserico. Community structure and interaction dynamics through the lens of quotes. *arXiv preprint*, 2016.