A Peer's-Eye View: Network Term Clouds in a Peer-to-Peer System

Raynor Vliegendhart
R.Vliegendhart@tudelft.nl

Martha Larson
M.A.Larson@tudelft.nl

Christoph Kofler
C.Kofler@tudelft.nl

Johan Pouwelse
J.A.Pouwelse@tudelft.nl

Delft University of Technology, Mekelweg 4, 2628 CD Delft, The Netherlands

ABSTRACT

We investigate term clouds that represent the content available in a peer-to-peer (P2P) network. Such network term clouds are non-trivial to generate in distributed settings. Our term cloud generator was implemented and released in Tribler—a widely-used, server-free P2P system—to support users in understanding the sorts of content available. Our evaluation and analysis focuses on three aspects of the clouds: *coverage*, *usefulness* and *accumulation speed*. A live experiment demonstrates that individual peers accumulate substantial network-level information, indicating good coverage of the overall content of the system. The results of a user study carried out on a crowdsourcing platform confirm the usefulness of clouds, showing that they succeed in conveying to users information on the type of content available in the network. An analysis of five example peers reveals that accumulation speeds of terms at new peers can support the development of a semantically diverse term set quickly after a cold start. This work represents the first investigation of term clouds in a live, 100% server-free P2P setting.

Categories and Subject Descriptors:
H.3 [Information Storage and Retrieval]: H.3.3 Information Search and Retrieval—*Search process*; H.3.4 Systems and Software—*Distributed systems*

General Terms: Performance, Experimentation

Keywords: P2P, term cloud, gossip protocol, user study

1. INTRODUCTION

New users encountering a search system can search more effectively if they have appropriate expectations of the sort of content that can be found in the system. Tribler is a real-world peer-to-peer (P2P) file-sharing system (downloadable from http://www.tribler.org) that offers a search functionality [7]. We developed and implemented a term cloud generator in order to promote successful searches by providing users with an impression of the types of content available in the system. Informal observation of user interaction patterns suggests that users having more experience with the Tribler system formulate a greater number of successful queries. The term clouds are intended to provide a quicker substitute for system interaction experience. If the clouds support users in understanding which information needs Tribler can fulfill, it can be expected that their queries better match the content of the system, leading, in the long term, to higher satisfaction and better user retention rates.

The term clouds are generated using the frequency counts of terms extracted from the names of files within the network that are accumulated at an individual peer by way of the underlying process used to exchange information among peers. This paper investigates the question of whether effective term clouds reflecting overall network content can be created in a distributed environment. We focus on three aspects: *coverage*, *usefulness* and *accumulation speed*. Note that this focus excludes investigation of cloud animation. Here, we simply state that animation switches cloud views at regular intervals to give the user the impression of the scope and dynamic development of the content of the system. Analysis of use pattern statistics and long-term impact on the uptake of Tribler are also left for future work.

In a completely distributed environment such as Tribler, building a network term cloud is non-trivial. Within the network, content is stored not on a central server with a 'bird's-eye' view, but rather at the individual peers. An individual peer can receive information about content at other peers only by communicating with its direct neighbors. In other words, in an environment that is 100% server free, the only view of the content collection that is available is a 'peer's-eye' view. In order for term clouds to be useful, the communication between peers must provide fast and high-coverage information about the content in the network. The key contribution of this paper is to demonstrate that a server-free architecture does not prevent peers from generating clouds that provide a global overview and are helpful for users.

After presenting background and related work, we report results of a live discovery experiment investigating cloud *coverage*, i.e., how well 'peer's-eye' clouds reflect network-level content. Then, we investigate the *usefulness* of the clouds, i.e., their ability to convey an impression of the content of the network to users, with a user study. Finally, we examine the *accumulation speed* of the clouds with a qualitative analysis of the cold start phase of example peers that reflects the experience of new users entering the network. We finish with our conclusion and outlook.

CIKM'11, October 24–28, 2011, Glasgow, Scotland, UK.

2. BACKGROUND AND RELATED WORK

The basic motivation for using term clouds to communicate the content of the system to users derives from work on tag clouds, which shows that clouds support browsing and discovery [8]. Our term clouds contain mixed uni- and bigram terms. This design choice is based on studies showing user preference for bigram or mixed clouds [4], which suggest that longer terms are easier for users to interpret. Since our aim is to investigate the viability of peer's-eye clouds in a P2P setting, we do not focus on the specific benefits derived from individual design characteristics of clouds, e.g., the use of terms vs. tags or the benefits of mixing terms in clouds.

Understanding our network term cloud requires careful distinction between 100% server free P2P systems and solutions with central dependencies. The presence of a central component in the design of a P2P system allows a significant simplification of the search set up, e.g., search in Napster using a central file index. In such a system, generation of term clouds is trivial. However, in a fully decentralized, i.e., 100% server free, P2P system such as Tribler, there is no central point that can, e.g., aggregate term frequencies and peers are required to propagate or search for information throughout the network. Depending on its network topology, a P2P system can use different communication protocols based on, e.g., flooding [5]. Gossip-based algorithms [2] provide the relative advantage of scalability and Tribler uses its own specific gossip protocol called Buddycast [6]. Through the exchange of periodic Buddycast messages, a peer discovers other peers and new content. Each message contains a list of live peers (divided into peers similar to the sending peer and peers that have been selected randomly), a download profile of the sending peer and a list of selected content hashes known by the sending peer. When a peer finds a peer with a similar download profile or an unknown content hash, it can connect to that peer or request the metadata of the file corresponding to that hash. Note that although the similarity relationship between connected peers is a distinguishing characteristic of Tribler, here, we exclude it from consideration in order to ensure our results achieve better generalization beyond Tribler to server-free P2P in general. If a peer downloads a file, it retrieves the file's content using the BitTorrent protocol [1].

Work closely related to our own is limited, and arguably effectively restricted to a single research effort, [3]. This work proposes an architecture for aggregation and representation of information resources that enables tagging in a P2P network with a Distributed Hash Table (DHT) topology. A DHT is a structured P2P network in which nodes and values are assigned a key through a hashing function and can be found using key-based routing. The basic challenge, that of information aggregation in a distributed environment, faced in [3] is shared with our work. However, our work differs in that Tribler is an unstructured P2P network, with greater flexibility and lower security risk. Further, here, we focus on term discovery, while [3] investigates maintaining frequency approximations of previously-known tags.

3. NETWORK TERM CLOUDS

Network term clouds are created by extracting terms from the filenames and accumulating raw counts. Peers acquire filenames from neighbors through torrent files and in turn pass these torrent files along again. The collecting of torrent files is driven by the Buddycast protocol, which allows peers to discover new content as described in the previous section. A torrent file contains metadata required for the BitTorrent protocol to download an actual file [1]. The metadata includes the filename, size and integrity hashes. From a filename both unigram and bigram terms are extracted. The unigram terms are obtained by tokenizing the filename using non-alphanumeric characters as delimiters. English stopwords and unigrams shorter than three characters are ignored. Bigram terms are constructed by joining the first two unigrams extracted from a filename, based on the assumption that the most important unigrams are at the beginning of a filename. The extracted unigram and bigram terms are displayed together in a single mixed cloud.

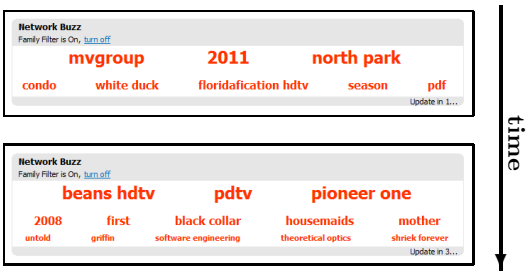

Figure 1: Network term cloud: peer's-eye view (two snapshots of the animated cloud in Tribler)

The network term cloud in Tribler is illustrated in Figure 1, which pictures two frames of the cloud, which is animated. Each frame of the animated term cloud shows for five seconds a random sample of 13 terms from high, medium and low frequency levels represented on three different tiers. The design decision to include three levels of frequency enhances the user's impression of the network content as evolving, since new terms are low frequency and would be completely excluded from the cloud, had frequency been the sole criterion for inclusion in the cloud. We restrict ourselves here mentioning the tiers, but do not investigate them further in the current work. The network term cloud can be observed live by downloading and installing Tribler; as an alternative we provide a demonstration video: http://youtu.be/hZeQlf5V8tA.

4. LIVE TERM DISCOVERY EXPERIMENT

The first aspect of the network term cloud we investigate is *coverage*. We carry out a live experiment within the Tribler network whose aim is to discover whether peer's-eye views of the network are mutually exclusive or whether the coverage of terms accumulated at a peer is substantial enough to support the generation of a cloud representing the overall content of the network. We acquired filenames at each of a pool of 30 peers under our control during their normal operation within the P2P system over an extended period of time (ca. 671 hours) and extracted terms from them. The 30 peers were started in succession on a single machine, joined the network, and only executed Buddycast to discover new peers and new content. Our peers did not initiate any downloads and did not build up a download profile. In this way, we ensured that they only connected to random and not semantically similar peers, as mentioned above. During the whole experiment they did not leave the system.

In Figure 2, discovery of terms from filenames is illustrated over time for the first four hours of the experiment.

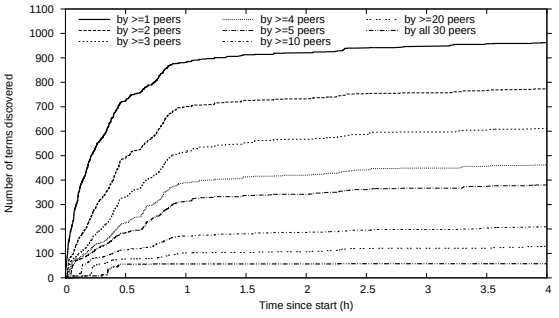

Figure 2: Peer-level discovery of terms over time

Each line represents the intersection of the number of terms discovered by a given number of peers, with the bottom line showing terms discovered by all peers. This figure yields several important insights. First, peers are far from discovering mutually exclusive term sets in the network, but rather a large overlap can be seen between the term sets that accumulate at the individual peers. Recall, that it is not possible to collect a complete global view of network content and that we approximate this view using the pool of peers under our control. In particular, we assume that terms discovered by at least one of the peers are representative of the overall content of the system. Figure 2 shows that peers in our pool discover a substantial proportion of the global term set. Second, the system does not reach a steady state, but rather the number of terms discovered by a single peer keeps growing and the other peers never converge with respect to the composition of their term sets. The growth reflects the constant entry of new content into the system. Although only the first four hours are pictured in Figure 2, the parallel growth trend is displayed during the entire collection interval of ca. 671 hours. Note that the flatness of the lowest lines in Figure 2 before ca. 0.4 hours can be attributed to the sequential startup of the peers. In order to get better insight on the early startup phase of peers, we will return later to investigate term accumulation immediately after the cold start of a peer.

We carried out an additional analysis to investigate the nature of those terms that are discovered by some peers, and thus assumed to belong to the global view, but not by others. In particular, we are interested in determining whether individual peers are likely to miss high frequency terms, under the assumption that such terms are the most important for characterizing the network-level collection. Figure 3 plots the probability that a term will fail to be discovered by one of the peers in our pool against the global frequency of that term estimated using the peer pool at two time-points in the life of a peer: a relatively immature (4 hours) and a mature (24 hours) stage. The exact times were chosen according to Tribler-specific considerations: four hours is the smallest resolution with which we can observe peers entering and leaving the network and 24 hours is the point at which Tribler considers the startup phase to have ended and switches messaging to a lower rate. We use Figure 3 to draw important general conclusions. First, it can be seen that the terms the most likely to be missed are in general low frequency terms. For high frequency terms, the miss probability approaches zero. Second, although the mature peer has fewer terms with high miss probabilities, the difference with the immature peer is not staggering, suggesting that the discovery process is already effective early in the life of a peer. In sum, the 'peer's-eye' view does not vary radically from peer to peer and does well in approximating the 'bird's-eye' view, generally missing a relatively restricted set of lower frequency terms.

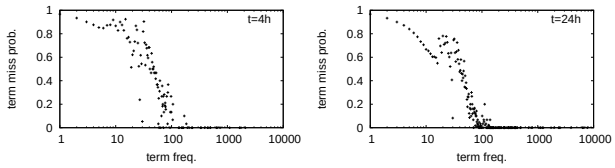

Figure 3: Term discovery failure vs. term frequency

5. USER STUDY

The second aspect of the network term cloud we investigate is *usefulness*. We performed a user study to investigate the question of whether term clouds can act as a surrogate for user experience with Tribler. We measure understanding of the collection in terms of the ability of a user to classify a file as either available or not available in Tribler. Informal observation of user interaction patterns provides evidence that a preponderance of Tribler queries correspond to known item search needs. We assume that the ability of a user to predict the availability of a file in Tribler reflects the ability to formulate successful searches.

We carried out experiments using Amazon Mechanical Turk (http://www.mturk.com), a crowdsourcing platform providing access to a pool of workers. Details of the study design and set up, including the crowdsourcing quality control mechanism to ensure serious user study participants, are described in [9]. The study investigates two conditions: `no-cloud`, in which the subject is presented with a mockup and a basic description of the system, and `with-cloud`, in which the subject is additionally presented with a series of five term clouds, such as they would be viewed (in animated sequence) in the system. The clouds were selected randomly using the set of terms that had been discovered after four hours by a typical peer chosen from our peer pool.

The file list we ask the subjects of the user study to classify consists of 100 filenames, half drawn from the Tribler system and half fake. The fake filenames were generated to represent types of content that were chosen by a panel of 'expert' users with extensive Tribler experience. They correspond to five categories clearly not represented in the Tribler system: 'home videos', 'news' and 'how-to videos', 'commercials' and 'sports'. The real filenames represented types of potentially findable content: 'TV', 'movies', 'music', 'software' and 'books'. The basis of the fake filenames were titles of existing videos from popular websites and titles of news items. The titles were modified to resemble the filenames of the real files, such that the difference between the two was disguised. This was done by adding plausible group names, format extensions and format specifications to the title.

In total, 184 workers took part in the user study, making a total of 4000 classification decisions divided over the 100 filenames. Subjects were first offered the `no-cloud` and then the `with-cloud` condition; they could participate in one or both conditions—offering this option effectively gave us access to more subjects. The filenames in each condition were mutually exclusive, but chosen to be comparable.

Table 1: User filename prediction (%correct)

	no-cloud	with-cloud
real	57.0%	63.1%
TV	66.3%	63.8%
movies	59.5%	69.5%†
music	59.5%	65.5%
software	50.5%	63.5%†
books	45.0%	55.0%†
fake	51.9%	52.1%
home videos	53.5%	49.5%
news	61.5%	61.5%
how-to	55.5%	61.0%
commercials	51.7%	42.5%
sports	33.8%	45.5%†
all	54.5%	57.6%†

† *Statistically significant improvement, Pearson's* χ^2 *test* ($\alpha = 0.05$)

Accuracy in the `no-cloud` condition was 54.5% and rose to 57.6% in the `with-cloud` condition (Table 1), a significant improvement according to Pearson's χ^2 test ($\alpha = 0.05$). The above-random performance in the `no-cloud` condition reflects assumptions that the users make about the content of the system from their prior file-sharing experience and the short description of the task. Out of 92 workers, 28 explicitly referred to the cloud when they were asked to justify one of their classification decisions, confirming that the clouds were consulted. Statistically significant improvements of the `with-cloud` over the `no-cloud` condition were observed in four categories: 'movies', 'software', 'books' and 'sports'. These results suggest that users do indeed gain an impression of the system content via the term cloud. The effect measured here is subtle, however, combined with aspects not yet studied here (e.g., use of cloud to support browsing, volume of user clicks) has a potential to increase user satisfaction with the system. Gains in specific categories less conventionally associated with file sharing (e.g., books) make term clouds particularly valuable for our P2P system.

6. COLD START ANALYSIS

The final aspect of the network term cloud we investigate is *accumulation speed*. We examine the cold start phase of five example peers from our pool with the aim of gaining an impression of the potential of very young clouds to be helpful to users, an issue assumed to be important for retention of new Tribler users. Our analysis procedure is based on the insight that it is not necessary for two clouds to be exactly identical in order to be equally useful to the user. For this reason, we concentrate not on the identities of cloud terms, but rather on their semantic diversity. We take diversity to reflect the ability of the clouds to convey an impression of the variety and scope of the content available in the network. We assume that certain terms in the cloud trigger users to infer the presence of certain types of content in the system. In particular, we focus on the categories of content known to exist in the system and used in the user study. For each of the terms in example clouds from our five peers, we make a best guess on which category it might reflect. In the case of TV vs. movies, it is often difficult to make a single best guess and in this case we label the term with a combined TV/movie category. Figure 4 shows the distribution of the terms over the categories in the clouds at four points along the life of the peer (2min, 10min, 4h and 24h). Between 2min and 10min clouds become mature enough to contain a full set of 13 terms. Although not all categories are present in the youngest (2min) clouds, the diversity is still good. Further, clouds at peers older than 10min are not radically more diverse, suggesting that very young clouds can be just as effective as mature clouds.

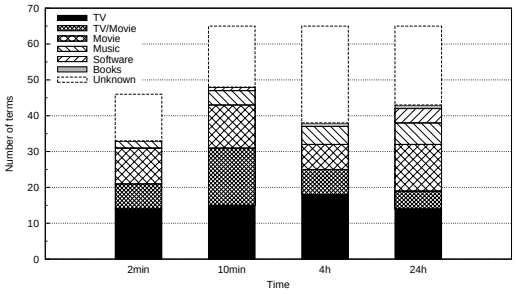

Figure 4: Semantic diversity of term clouds

7. CONCLUSION AND OUTLOOK

We have demonstrated that information aggregated at a single peer within a distributed system is adequate to support generation of term clouds that provide a user with useful information about the content of the system. Since the Tribler client attracts heavy use—it has been downloaded more than 800,000 times within the last five years—even a small or modest improvement in users' understanding of the system has the potential to lead to a large impact in terms of improved query success and better user experience. Further, communication of an impression of the global content of a P2P network to users is critical as P2P moves into new domains, since it helps users to quickly shake outdated assumptions about the nature of the items being shared by file sharing. Future work will involve analysis of the click behavior of users interacting with the term cloud and will shed light on the details of cloud use, particularly on the ability of clouds to support browsing and discovery and to improve the retention of new users of the system.

8. ACKNOWLEDGMENTS

The research leading to these results has received funding from the European Commission's 7th Framework Programme under grant agreement № 216444 (PetaMedia).

9. REFERENCES

[1] B. Cohen. Incentives build robustness in BitTorrent. In *Workshop on Economics of Peer-to-Peer systems*, 2003.

[2] P. Costa et al. Exploring the interdisciplinary connections of gossip-based systems. *ACM SIGOPS Operating Systems Review*, 41(5):51–60, 2007.

[3] O. Görlitz, S. Sizov, and S. Staab. PINTS: Peer-to-Peer Infrastructure for Tagging Systems. In *International Conference on Peer-to-Peer Systems*. USENIX, 2008.

[4] R. Kaptein, D. Hiemstra, and J. Kamps. How different are language models and word clouds? In *ECIR 2010*, volume 5993 of *LNCS*, pages 556–568, Berlin, March 2010.

[5] E. K. Lua et al. A survey and comparison of peer-to-peer overlay network schemes. *Communications Surveys & Tutorials, IEEE*, 7(2), 2005.

[6] J. Pouwelse et al. Buddycast: an operational peer-to-peer epidemic protocol stack. In *ASCI 2008*.

[7] J. Pouwelse et al. Tribler: A social-based peer-to-peer system. *Concurrency and Computation: Practice and Experience*, 20(2), 2008.

[8] J. Sinclair and M. Cardew-Hall. The folksonomy tag cloud: when is it useful? *J Inf Sci*, 34(1), 2008.

[9] R. Vliegendhart, M. Larson, C. Kofler, C. Eickhoff, and J. Pouwelse. Investigating Factors Influencing Crowdsourcing Tasks with High Imaginative Load. In *WSDM'11 Workshop on Crowdsourcing for Search and Data Mining*, February 2011.