Impact of Socially Based Demand on the Efficiency of Caching Strategy

Buster O. Holzbauer and Boleslaw K. Szymanski
Department of Computer Science
Rensselaer Polytechnic Institute
Troy, NY USA
holzbh@cs.rpi.edu, szymansk@cs.rpi.edu

Eyuphan Bulut
Cisco Systems
2300 President George Bush Highway
Richardson, TX USA
ebulut@cisco.com

Abstract—**Routers are often faced with a variety of users throughout a day. These users make requests, of which the most bandwidth consuming and least delay tolerant are requests for internet content including videos. To save cost of uploading all requested data, routers use content-aware caching schemes. The efficiency of such schemes is dependent on router ability to predict the use of the currently requested content. At the same time, research on social networks indicates that users with social ties have the correlated mobility patterns and the correlated interests in requested content. In this paper, we demonstrate that these correlations between mobility and between interests can be used to improve efficiency of content-based caching network caching by taking into account social ties between users visiting a router. To this end, we built a model for router visits and user demand based on social relations. We then used this model to create a content-aware caching scheme called "SOcial Content Caching Scheme (SOCCS)". Our results demonstrate that by using information about frequency of visit by members of different social communities combined with topic tags on the content requested by each community, the performance of routers can be increased remarkably compared to their performance with random caching. Moreover, such socially based caching is robust to changes in many parameters of our model that affect visits by mobile nodes, the tags of content that they request, and how history is incorporated into caching decisions.**

Keywords—*cooperative caching; social networks; demand model*

I. Introduction

In recent years the number of portable devices that can use content from the Internet, such as smartphones, laptops and tablets, has increased tremendously such that they are part of every day life for many individuals. As a result, hardware that can pull information from the Internet, like WiFi routers, are increasingly utilized. This happens not just in homes, but at workplaces, hotspots at transit hubs, coffee shops, and many other locations as well. Even mass transportation such as buses and trains may have available WiFi connections.

Due to the increasing demand on limited resources, such as bandwidth, and the need to have "reasonably fast" response times to customer requests, it is important that routers do their best to minimize bandwidth usage dedicated to Internet traffic. By installing a hardware cache, a router can retain media or other data that is requested repeatedly, and avoid having to download it again on subsequent requests. Unfortunately, space and budget constraints can significantly restrict cache sizes, and even with a relatively large cache, it is impossible to cache every request that will arrive at a router. For that reason we believe that developing a caching algorithm that can make informed decisions about what to cache can lead to better router performance.

The applications that are bandwidth heavy and delay tolerant benefit the most from caching, hence in this paper we focus on demand for video content. This is further justified by the popularity of Netflix and YouTube sites [1]. According to Cisco Visual Networking Index (VNI) [2] it is forecasted that two-thirds of the world's mobile data traffic will be video by 2017. In addition, many social networking platforms allow for sharing of video from external sites, or provide video services such as Vine [3] increasing the correlation between video demand originating within the same community. Because of our focus, we do not consider node-to-node communication, e-mails, or many other types of traffic that a router may see. Additionally, such focus allows us to concentrate on the idea of using content-aware caching, where a small amount of information about the data being requested can provide performance increases.

Communities are the basic structures in sociology in general and in social networks in particular. They have been intensively researched for more than a half of the century [4]. In sociology, community usually refers to a social unit whose members share common values and the identity of the members as well as their degree of cohesiveness depend on individuals' social and cognitive factors such as beliefs, preferences, or needs. In our approach we focus on the cohesiveness of preferences or needs which naturally applies to internet content, including videos. What is interesting in our approach that we do not need to discover communities explicitly, it is sufficient that routers observe the visiting nodes preferences. It is however essential that the video demand model is built with the assumption that the communities have correlated interests, something that has been confirmed by traditional sociological research [4], as well as by the observation of the followee-follower networks in Twitter [5].

The rest of the paper is organized as follows. In Section II we discuss related work. In Section III we describe the models we use to build our system, SOcial Content Caching Scheme (SOCCS). Section IV describes simulations done for validation and the results that follow from them. Finally in Section V we conclude by highlighting our findings and describing opportunities for future work.

II. Related Work

Due to the importance of mobility in many systems, such as participatory systems, delay-tolerant networking, and the application discussed in this study, there is a large body of work on modeling mobility and data access. Some approaches are trace-based [6], with traces such as the MIT Reality Mining set [7] and Haggle dataset [8]. Trace-based approaches are beneficial because they are derived from real data with human participants. They can be utilized for better prediction of node relations in future and for better routing decisions in DTNs [9], [10], [11]. However, they are often limited in both time span and in scope, which can lead to a biased population and traces that may not be suitable for general use.

Other approaches are strictly synthetic, and may be random, probabilistic, or even involve multiple nodes [12]. Many of these synthetic models can produce explicit locations, or require knowledge about prior locations and about the nature of mobile nodes to compute the next location of nodes.

In our approach, we focus on human mobility, as we are interested in caching for members of social communities. As established in [13], human mobility is highly predictable, even without considering social ties. Regardless of whether the radius of a node's travels is small or large (i.e. for a variety of travel distributions), regularity and subsequently predictability of location remains high. Additionally, their work shows that regularity of an individual's mobility is not affected by weekend vs weekday activity, meaning the home-work-entertainment pattern does not affect predictability. We use a concept of day workers and evening entertainment communities in our modeling, but primarily use it for the purpose of demand generation. In fact, there is no correlation in our system between nodes or mobility history in the day and evening communities - a node may actually be a member of both, but the system has no memory of the identity across time spans. The work by Song et al. suggests that the patterns in the two periods can be examined separately without compromising the validity of our mobility model.

However, the study done by Song et al. does not provide insight about social communities. Using location based social media, authors in [14] showed that mobility of members of the same community is highly correlated. Nguyen et al. developed a Markov model for each user based on their Gowalla check-ins, which is used in their Friendship Mobility Model (FMM). The FMM's visiting space is limited to locations the node has checked into using a location-based social service, Gowalla. While we do not attach any semantic meaning to visiting a particular router, it could be viewed that each router has a particular significance, such as an office building, shopping center, apartment complex, etc. The FMM is driven by real data which inherently creates affinities to particular locations. In contrast, in our work we still enforce an idea of affinities, but they are developed non-deterministically based on previous decisions of the node and its community. Nguyen et al. [14] present the important finding that there is a relationship between friendship and distance. This is important to our work because it means that given users in a community (i.e. friends), we can expect that they must move in such a way that they are likely to be close to each other.

Data access is also an important issue in Delay Tolerant Networks (DTNs), and a common technique to improve the performance of data access is cooperative caching. In [15], the authors proposed DAC, a novel caching protocol adaptive to the challenging environment of DTNs. In this protocol, they relay on social community structure to overcome a challenge of the dynamically changing network topology in DTNs. Using a new centrality metric, the approach evaluates the caching capability of each node within a community, to determine where to cache. Another important factor considered is contact duration limitation that is used to derive an adaptive caching bound for each mobile node. Using extensive trace-driven simulations, the authors show that their cooperative caching protocol can significantly improve the performance of data access in DTNs.

Moreover, in [16], the authors further showed that members of the same community tend to stay in the same locations together. This supports the assumption of our approach that the visiting events at a router will have high correlation among members of the same community. As a direct result, a router that observes a node belonging to a particular community can expect to see more nodes from the same community in the near future. This expectation drives our decision to use a correlated mobility model in our system design. By using a location of interest based system, we avoid the need for the routers to know exact positions of the nodes that interact with them. Locations of interest in our cases are routers at notable locations, this is similar to the idea of locations labeled with a checkin name in the Gowalla system studied by Nguyen et al [14].

The concept of caching data is not new. Store and forward based approaches have long been a part of communications and delay-tolerant networking research, and these require space to store, or cache data. While our intended use of caches is for repeated data instead of replication and connectivity, the ideas of choosing which packets to cache and attempting to predict demand are common to both uses. There are already some studies done in literature regarding community estimation and caching based on community-based metrics [17] and the impact of the contact duration limitation on cooperative caching [18]. However, our approach does not use user nodes as part of the delivery system and does not utilize information about contacts between user nodes.

Finally, an existing idea is considering content labels, which has already been applied to studying the differences in life cycles of videos on YouTube [19]. The idea has also been applied to Twitter, with the result that an information-based structure existed separate from the connectivity (follower-followee) structure, and that furthermore, a genome of topics and users can be constructed through analysis of a dataset [5]. The relationship between communities and content is of particular value to the work in this paper.

III. Models

To examine the effects of content-aware caching on network performance, we designed SOCCS based on three components: a correlated mobility model, a caching metric, and a demand model. These three aspects are independent and any one part can be altered without having to change the other two. Within our system, the world is modeled as a number of agents
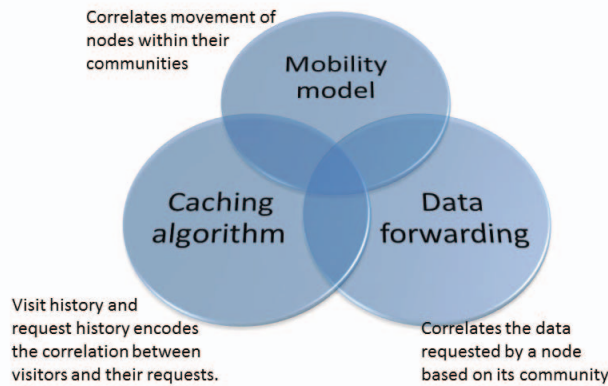
Fig. 1: The three components of our system. Each of the models is designed independently, but they interact closely during execution.

which travel and request data from stationary routers. The data is tagged only with a *topic*, such as Politics or Business. Our experiments are done to illustrate the effects of caching with only low-resolution data, but we expect the benefit would be greater given higher-resolution tags such as "music celebrities," "foreign policy," or "basketball news." In Figure 1, the basic models in our system are shown along with a brief summary of each one.

A. Correlated Mobility

One of our assumptions is that communities have correlated mobility patterns (which is also verified by previous work [20]). This means that given only a limited sample set of visitors, a router can assume that a similar mix of communities will be seen in the near future. As time changes, communities may alter their behavior or frequency. For example, consider a router in an office building that observes a community representing day workers. The bulk of workers will arrive between 8am and 9am, and leave between 4pm and 5pm. As a result, the number of visitors from this community will climb during the 8am-9am period, and likely have little change until 4pm-5pm. When the router begins to see the increase in the morning, it can predict that the day worker community will continue to have a presence in its range. Once the number of visitors stabilizes, the router can predict that while there will not be further significant growth in the near future, the number of visits should continue to be similar to the current estimate. As the day workers start to leave, the router can predict that there will be a decrease in visits. While our model (discussed in Section III-C) does not look at the current trend in changes to predict how future visits will change, it does make use of the assumption of community mobility to keep a smoothed estimate of future arrivals.

Given an aging coefficient α, number of visits by node j at router k at instant t v_{jkt}, and initial seeds for $\forall i \epsilon N, k \epsilon R, M_{ik0}, N$, where N is the set of all nodes and R is the set of all routers, the probability of a node in community c visiting router k at time t, P_{kct} is found by the following equations:

$$P_{kct} = \frac{R_{kct}}{R_{ct}} \quad (1)$$

$$R_{ct} = \sum_{k \epsilon R} R_{kct} \quad (2)$$

$$R_{kct} = \sum_{j \epsilon c} \frac{M_{jkt}}{M_{kt}} \quad (3)$$

$$M_{kt} = \sum_{i \epsilon N} M_{ikt} \quad (4)$$

$$M_{ikt} = \alpha * M_{ik(t-1)} + (1-\alpha) * A_{ikt} \quad (5)$$

$$A_{ikt} = \sum_{l \epsilon (t-1,t]} v_{ikl} \quad (6)$$

In the above equations, A_{ikt} is simply the count of visits by node i at router k in the current aging window. M_{ikt} is an estimate using α to smoothly incorporate visiting data from prior windows. R_{kct} is the weighted frequency of visits from nodes in community c at router k in the current aging window, including information from the past. Finally P_{kct} is simply the weighted frequency of visits from any node in community c to router k normalized by the total of all weighted frequencies across all routers for community c.

Due to the recursive nature of the above probabilities, as a particular router is visited by members of a community, it becomes more likely that it will be selected by other members in the community. By altering α, the strength of this effect can be adjusted. In addition, setting the initial M_{ik0} values allows experiments to determine the initial router preference. The values, and not just the proportions of values, matter. Changing the length of the window, T, also alters behavior, since the α-based aging is only applied at the end of every window.

B. Demand Model

We assume that the interests of nodes in a community are related, but not necessarily identical. This allows us to model the tastes of individual nodes by drawing from a common distribution in the community. Motivated by existing research that uses the idea of genotyping [5] to categorize data and users, we use the same 5 categories as in this reference: Sports, Politics, Business, SciTech, and Celebrities. Each community contains a distribution for each topic, and individual nodes draw from the distribution and normalize the values to sum to 1. These values then become the probabilities of a node requesting data from a specific topic.

In our case, we consider video demand, and abstract a particular request to always being the whole video (or part of video), to keep things simple. In reality, videos may be of varying sizes and nodes may only request a part of the video, or even a single clip. These considerations could affect the efficacy of caching, but our current approach is to illustrate that content-aware caching has an effect even with a low-detail scenario. We believe that some videos will be more popular in a given category, thus, we use a Zipf distribution [21] to indicate their relative probabilities. In our experiments, we limit the number of videos to 5, since even with just

Fig. 2: Effect of α on Cache Miss Rate
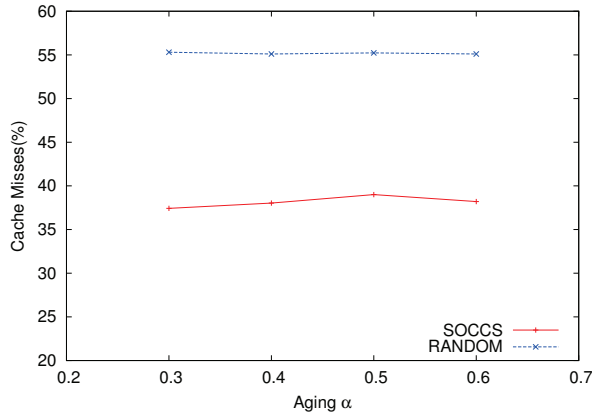
Fig. 3: Effect of Demand Estimate Seed on Cache Miss Rate

5 elements, the fifth number from a Zipf distribution has low probability of being chosen. The least popular video or two for a topic can be viewed as a meta-video that actually represents many less popular videos. While we elect to use a Zipf distribution for each topic, resulting in seemingly the same probabilities if the number of videos are the same, other research shows that videos can have different demand over a life cycle based on content type [19]. However, this is also the case in our model, because the final frequency of video demand is the product of the probability defined by Zipf distribution and the frequency of visits by the members of communities demanding the corresponding topic.

C. Caching Model

SOCCS only requires collecting the demands of nodes that are currently making requests from a router. Each router uses the same equations as described in Section III-A, but instead of v_{ikl} referring to node i on router k at time l, it instead refers to topic i. Similarly, all other equations refer to topics instead of nodes, but they require no updates since everything ultimately uses the definition of v_{ikl}.

We also assume that the caching model has some estimate of the relative popularity of videos in a topic. While in reality such knowledge cannot be current and perfect, it is unlikely that the vast majority of videos will experience a sudden and drastic change in demand. By having time to build a history of video requests, a router could have a good estimate of demand for existing videos. In our experimentation we looked at a daily cycle, so a window of 24 hours would likely be sufficient to obtain an estimate. Using this method, a router cannot anticipate demand for newly created videos or content that suddenly becomes popular. These details are outside the scope of our current work, and are factors that are not present in our demand model described in Section III-B.

When a request is made for new data from a router, it decides whether or not to cache a copy of the data. If there is room in its cache, which is of limited size β, then the router automatically caches the data. Otherwise, it computes the score of the data in its cache and the new data, and replaces the lowest scoring item with the new data only if the new data has a higher score. The score $S_{dc}(i)$ at time t for data d belonging
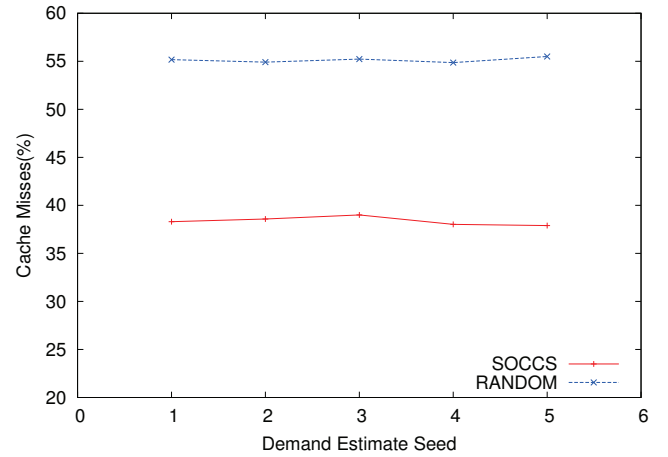
to topic i, requested by a member of community c is calculated as follows:

$$S_{dc}(i) = P_{ict} * P_{di} \quad (7)$$

P_{ict} is the probability that a member of community c will request data belonging to topic i on the current router. P_{di} is the probability that video d is requested, given that a video of topic i is being requested. Note that while the probability of selecting a particular topic depends on the community the request came from, the probability of selecting a given video is the same across communities.

IV. SIMULATION

We created a simple simulation to study the three models described in Section III. Since the goal was to illustrate that our caching model can help reduce the number of cache misses, we compare our caching algorithm's performance against random selection. In random selection, if there is no room to cache new data, a random element from the cache is replaced by the new data.

The experiments were performed by placing three static routers and then running a system of 1,800 nodes over the course of a 17 hour period from 8 AM. to 1 AM. Since in the real world it is unlikely that all the nodes from a community would visit any given router, we introduced 300 nodes per community with the expectation that these nodes represented only a part of their respective communities. Communities 1-3 were day communities, while communities 4-6 were evening communities.

Nodes belonging to day communities started visiting routers randomly with their times drawn from a uniform distribution between 8 AM. and 9 AM. They then stopped visiting at time drawn from a uniform distribution between 4 PM and 5PM. At the same time, the evening communities started randomly at a time drawn from the uniform distribution from 4PM and 5PM, and ended at a time drawn from a uniform distribution between 12 AM and 1 AM. This allowed for a gradual introduction and departure of communities, as well as a mixed composition between 4PM and 5PM.
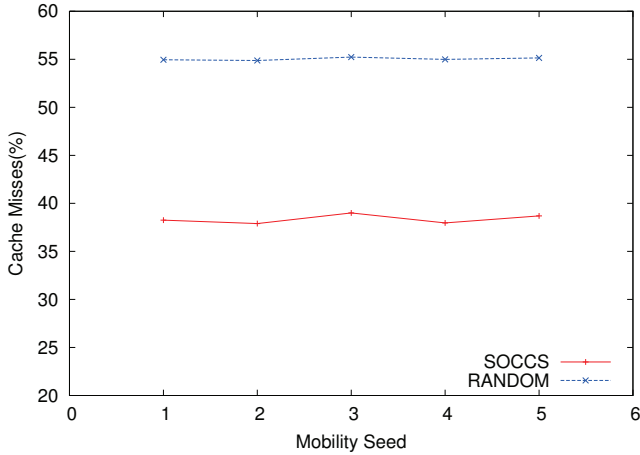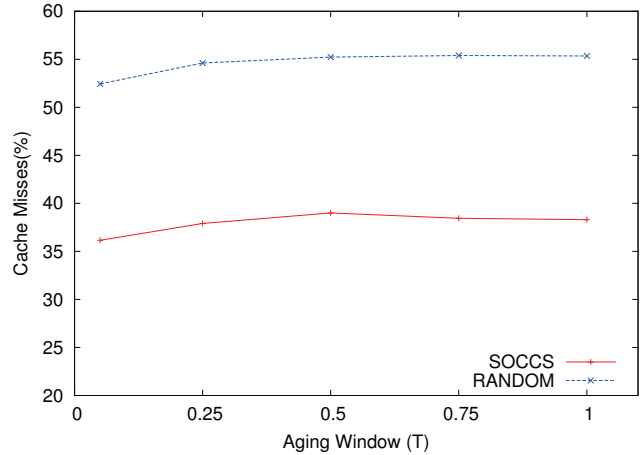
Fig. 4: Effect of Mobility Seed on Cache Miss Rate

Fig. 5: Effect of Aging Window T on Cache Miss Rate

The time between each visit was done by using an exponential distribution with mean inter-arrival time ($\frac{1}{\lambda}$) to introduce randomness. Each node drew its demands from five Gaussian distributions, one per topic, then normalized the demands so the total demand summed to 1. The Gaussian parameters for each community are listed in Table I.

TABLE I: Community Genotype Distributions

Community	Sports	SciTech	Business	Politics	Celebrities
1 (μ,σ)	3,1	3,1	30,5	60,10	3,1
2 (μ,σ)	60,10	3,1	30,5	3,1	30,5
3 (μ,σ)	20,3	20,3	20,3	20,3	20,3
4 (μ,σ)	3,1	3,1	30,5	3,1	60,10
5 (μ,σ)	60,10	3,1	3,1	3,1	30,5
6 (μ,σ)	20,3	20,3	20,3	20,3	20,3

Fig. 6: Effect of Cache Size on Cache Miss Rate

While we examined a variety of values for T, α, and the values of M_{ik0} for both demand estimation and correlated mobility, we found that none of these factors affected the cache hit rate by more than 1-2%. These results are shown in Figures 2- 5, which use a cache size of 8. As the plots in all these four figures show, the proposed content-aware algorithm (SOCCS) has cache misses around 60%-65% of the cache misses in random algorithm (RANDOM). If we compare the performances in terms of the cache hit ratios, we see that the cache hit ratio with SOCCS algorithm is around 40% higher than the cache hit ratio with RANDOM algorithm.

On the other hand, as expected, cache size has a significant impact on the performance of both algorithms. In Figure 6, we show the impact of cache size on the cache misses of both algorithms while all other parameters are fixed. It is apparent that the RANDOM algorithm always causes significantly more bandwidth usage (since content is not available in cache) than the SOCCS algorithm for a given cache size. This is expected, since our model suggests that many members of the same community will likely visit a router (due to correlated mobility) and communities share similar demands. Cache size still plays an important part, since with more cache slots, a router can afford to select less popular items to cache
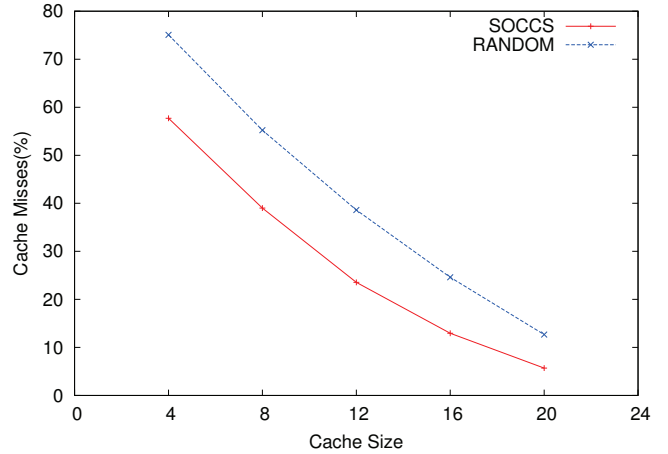
and still have room for popular items. Cache misses can never drop to 0%, since data cannot be pre-cached and must be downloaded at least once. Moreover, the percentage of saving in bandwidth usage with SOCCS algorithm compared to RANDOM algorithm changes depending on the cache size. As the cache size increases, the cache misses for both algorithms decrease, thus, bandwidth usage decreases. As a result, with increasing cache size, the relative bandwidth saving achieved by SOCCS algorithm compared to bandwidth usage with RANDOM algorithm increases.

V. CONCLUSION

Inspired by the problem of caching with limited space while serving video content to users, we adopted a content-based approach to generate a simple caching metric, SOCCS. We then constructed a simulation based on the assumptions that mobility was correlated among nodes in a community, and that there was a correlation between genotypes of users in communities. Under these conditions, our caching algorithm could perform significantly better than random selection. Simulations also showed that both algorithms were relatively insensitive to

a variety of parameters, but caching was significantly impacted by changes in the router cache sizes.

An option for future work is to improve the models for mobility and demand, such as by using a trace-based model to provide a more realistic estimate. Using information from social networks such as Foursquare could also provide us with improved demand and mobility models. Another option is expanding our approach to a mixed situation where routers may also be nodes, and may be mobile. This would allow us to then let routers cache content-tagged data, but also perform more traditional DTN routing tasks. By having routers be user nodes, they could benefit from additional information such as knowing their communities and currently reachable nodes. However, this might come with extra challenges of finding communities in DTN environment for which we can benefit from many studies [22], [23], [9] reported in the literature on this subject. Additionally, while we have shown that our approach works within a simulated environment, validating our assumptions and our approach in the real world, even on a small scale, is also a task that we may consider.

Acknowledgement

References

[1] Youtube.com site info (alexa). http://www.alexa.com/siteinfo/youtube.com. Accessed: 2014-01-18.

[2] Cisco visual networking index: Global mobile data traffic forecast update, 20122017. http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-520862.html. Accessed: 2014-01-21.

[3] Vine: A new way to share video. https://blog.twitter.com/2013/vine-a-new-way-to-share-video. Accessed: 2014-01-18.

[4] R. E. Park, *Human communities: The city and human ecology*. Free Press, 1952.

[5] P. Bogdanov, M. Busch, J. Moehlis, A. K. Singh, and B. K. Szymanski, "The social media genome: modeling individual topic-specific behavior in social media," in *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. ACM, 2013, pp. 236–242.

[6] S. C. Geyik, E. Bulut, and B. K. Szymanski, "Grammatical inference for modeling mobility patterns in networks," *IEEE Trans. Mob. Comput.*, vol. 12, no. 11, pp. 2119–2131, 2013.

[7] N. Eagle and A. (Sandy) Pentland, "Reality mining: sensing complex social systems," *Personal Ubiquitous Comput.*, vol. 10, no. 4, pp. 255–268, Mar. 2006. [Online]. Available: http://dx.doi.org/10.1007/s00779-005-0046-3

[8] A european union funded project in situated and autonomic comm. www.haggleproject.org. Accessed: 2014-01-18.

[9] E. Bulut and B. K. Szymanski, "Exploiting friendship relations for efficient routing in mobile social networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 23, no. 12, pp. 2254–2265, 2012.

[10] E. M. Daly and M. Haahr, "Social network analysis for routing in disconnected delay-tolerant manets," in *Proceedings of the 8th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, ser. MobiHoc '07. New York, NY, USA: ACM, 2007, pp. 32–40. [Online]. Available: http://doi.acm.org/10.1145/1288107.1288113

[11] P. Hui, J. Crowcroft, and E. Yoneki, "Bubble rap: Social-based forwarding in delay tolerant networks," in *Proceedings of the 9th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, ser. MobiHoc '08. New York, NY, USA: ACM, 2008, pp. 241–250. [Online]. Available: http://doi.acm.org/10.1145/1374618.1374652

[12] T. Camp, J. Boleng, and V. Davies, "A survey of mobility models for ad hoc network research," *Wireless Communications and Mobile Computing*, vol. 2, no. 5, pp. 483–502, 2002. [Online]. Available: http://dx.doi.org/10.1002/wcm.72

[13] C. Song, Z. Qu, N. Blumm, and A.-L. Barabasi, "Limits of predictability in human mobility," *Science*, vol. 327, no. 5968, pp. 1018–1021, 2010.

[14] T. Nguyen and B. K. Szymanski, "Using location-based social networks to validate human mobility and relationships models," in *Proceedings of the 2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. ACM, 2012, pp. 1237–1253.

[15] X. Zhuo, Q. Li, G. Cao, Y. Dai, B. Szymanski, and T. La Porta, "Social-based cooperative caching in dtns: A contact duration aware approach," in *Proceedings of the 2011 IEEE 8th International Conference on Mobile Adhoc and Sensor Systems*. IEEE, 2011, pp. 92–100.

[16] T. Nguyen, M. Chen, and B. Szymanski, "Analyzing the proximity and interactions of friends in communities in gowalla," in *Proceedings of the 2013 IEEE 13th International Conference on Data Mining Workshops*, ser. ICDMW'2013. Los Alamitos, CA, USA: IEEE, 2013.

[17] W. Gao, G. Cao, T. La Porta, and J. Han, "On exploiting transient social contact patterns for data forwarding in delay-tolerant networks," *Mobile Computing, IEEE Transactions on*, vol. 12, no. 1, pp. 151–165, 2013.

[18] X. Zhuo, Q. Li, G. Cao, Y. Dai, B. K. Szymanski, and T. L. Porta, "Social-based cooperative caching in dtns: A contact duration aware approach." in *MASS*. IEEE, 2011, pp. 92–101. [Online]. Available: http://dblp.uni-trier.de/db/conf/mass/mass2011.html#ZhuoLCDSP11

[19] M. Cha, H. Kwak, P. Rodriguez, Y.-Y. Ahn, and S. Moon, "I tube, you tube, everybody tubes: Analyzing the world's largest user generated content video system," in *Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement*, ser. IMC '07. New York, NY, USA: ACM, 2007, pp. 1–14. [Online]. Available: http://doi.acm.org/10.1145/1298306.1298309

[20] E. Bulut, S. C. Geyik, and B. K. Szymanski, "Efficient routing in delay tolerant networks with correlated node mobility," in *MASS*, 2010, pp. 79–88.

[21] G. K. Zipf, *Human Behaviour and the Principle of Least Effort*. Addison-Wesley, 1949.

[22] F. Li and J. Wu, "Localcom: A community-based epidemic forwarding scheme in disruption-tolerant networks," in *Proceedings of the 6th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks*, ser. SECON'09. Piscataway, NJ, USA: IEEE Press, 2009, pp. 574–582. [Online]. Available: http://dl.acm.org/citation.cfm?id=1687299.1687364

[23] P. Hui, E. Yoneki, S. Y. Chan, and J. Crowcroft, "Distributed community detection in delay tolerant networks," in *Proceedings of 2Nd ACM/IEEE International Workshop on Mobility in the Evolving Internet Architecture*, ser. MobiArch '07. New York, NY, USA: ACM, 2007, pp. 7:1–7:8. [Online]. Available: http://doi.acm.org/10.1145/1366919.1366929