Posters

P1. Analysis of Content and Activity in Geo-Social Networks for Place Recommendations
Anastasios Noulas and Cecilia Mascolo (University of Cambridge, Computer Laboratory)

P2. Can We Exploit the Wisdom of Large Ad Hoc Crowds?
Simon Fleming, Ian Wakeman and Dan Chalmers (University of Sussex)

P3. Collaborative Privacy Policy Authoring in a Social Networking Context
Ryan Wishart (Imperial College London)

P4. Distance Matters: Geo-social Metrics for Online Social Networks
Salvatore Scellato, Cecilia Mascolo (University of Cambridge), Mirco Musolesi (University of St. Andrews) and Vito Latora (University of Catania)

P5. Empowering people to fight killer flu with mobile phones and social networking
Chris Watkins (Department of Computer Science, Royal Holloway)

P6. Growth mechanisms in continuously-observed networks: Communication in a Facebook-like community
Tore Opsahl (Imperial College Business School, Imperial College London)

P7. How social media helps the music industry: Tracking artist popularity, fan interactions and media consumption using Musicmetric
Trung Huynh, Gregory Mead, Jameel Syed, and Matthew Jeffery (Musicmetric - www.musicmetric.com)

P8. Open Innovation: how corporate R&D scientists capture and exploit external knowledge from their networks
Anne ter Wal, Paola Criscuolo, and Ammon Salter (Innovation Studies Centre, Imperial College Business School)

P9. Privacy concerns and social network routing
Iain Parris and Tristan Henderson (School of Computer Science, University of St Andrews)

P10. Small-world behavior in time-varying graphs
John Tang, Salvatore Scellato, Mirco Musolesi, Cecilia Mascolo (Computer Laboratory, University of Cambridge) and Vito Latora (Università di Catania)

P11. Studying Location Sharing on Social Networks with Mobile Experience Sampling
Fehmi Ben Abdesslem, Iain Parris and Tristan Henderson (School of Computer Science, University of St Andrews)

P12. Word-of-Mouth based Web Discovery: URL Propagation in Twitter
Jisun An (University of Cambridge), Tiago Rodrigues and FabrÌcio Benevenuto (Universidade Federal de Minas Gerais), Meeyoung Cha (Graduate School of Culture Technology, KAIST), and Krishna Gummadi (Max Planck Institute for Software Systems)

Analysis of Content and Activity in Geo-Social Networks for Place Recommendations

Anastasios Noulas
Computer Laboratory
University of Cambridge
name.surname@cl.cam.ac.uk

Cecilia Mascolo
Computer Laboratory
University of Cambridge
name.surname@cl.cam.ac.uk

Introduction While the area of Geo-Social networking (GSN) is growing rapidly, with the rates of user adoption and appearance of startups experiencing exponential growth during 2010, there are series of research questions that arise synchronously to this technological shift. We identify one of the primary ones to be the division of frameworks that will be able to analyse and model data sourced from those systems. The necessity for such direction primarily originates from the fact that the data generation process features properties that may require special attention. First, it is *quantity diverse* from the point of view that user participation in those systems may vary, for example at an individual level as one may forget to checkin at a place or may avoid it for privacy reasons. Quantity of data reception may also be across different geographical areas (e.g. cities) that have different levels of adoption. Second, GSN data is *quality diverse* as one may observe mobility (visits at exact locations) trajectories of users, activity (visits at types of places) trajectories or text. The latter, may be user communication content or tips and tags assigned to locations. Further in this document, we describe our first steps towards addressing some of the requirements in creating a layer towards the unification of GSN data that will facilitate the access of applications and services to it.

Modelling Geographical Areas We initially present a framework for the representation *geographical areas* by means of nearby points of interest. Given a geographic point p encoded through a (latitude,longitude) pair, we retrieve all nearby a locations l within radius r and their respective categories c. The constitution of the considered area in terms of places can be indicative of what a user could encounter in the area. Moreover, we incorporate social activity to our methodology, by observing the number of *checkins* at a location. As an experimental case that utilises the above representation, we apply a clustering algorithm over the city of London and we discover groups areas where similar activity patterns.

Topic-Modelling and Sentiment Analysis applied on Communication Content With respect to user generated content, such as text in Geo-social networks, we propose the application of two natural language processing based paradigms, each of which will serve a different purpose. Topic Modelling, applied on micro-blogging data featured in those systems, could provide information about what is a place or a general area about. In addition, sentiment analysis techniques could provide information about the emotional consensus of the user community across the geographic plane. We envision a system, that could employ this analysis in order to provide automated reviews on places and areas as people move by and discuss about them.

An eye to future applications Models of data generated in Geo-Social Networks could be useful to a wide range of applications and services. Urban planners could discover more about a city, since the dynamics of human activity on it are revealed as people checkin at places. Augmented reality applications and location based services could also benefit as new data recourses can be accessed at massive scales. In addition, social scientists could learn more on how social systems evolve as patterns of individuals and social interactions emerge. To close, we present an example of how GSN data could be used in an application context by considering a place recommendation scenario for tourists at two metropolitan cities.

Can We Exploit the Wisdom of Large Ad Hoc Crowds?

Simon Fleming, Ian Wakeman and Dan Chalmers
University of Sussex
{s.fleming, ianw, d.chalmers} @sussex.ac.uk

October 28, 2010

When people are joined together over communication networks, it is possible to ask questions and retrieve high quality answers using the wisdom of the crowd. But locating a suitable candidate for answering a given question within a large ad hoc crowd is non-trivial. We wish to source the best answers possible from the network, while at the same time controlling the levels of attention required from the crowd (a novel routing metric). We envisage the concept of a distributed question and answer service over ad hoc networks, focusing on fully decentralised methods and protocols to route questions towards members of the network who may be able to answer well. We are motivated by solving user privacy concerns, by allowing both question asking and answering to be plausibly deniable.

We thus define our application scenario: *Active network members may submit textual questions into the network at any time. Questions will use single hop routing tactics to jump between pairs of nodes aiming to find a suitable answerer. When a question reaches a node which is interested, it is recorded and dealt with as and when the user has time to do so. Generated answers will follow the same route that the original question took back to the originating node. The choice of tactics used for routing questions will adjust the resulting quantity and quality of answers and also the path lengths. We assume users require the highest possible answer quality while reducing the overhead in terms of user attention.*

Our poster will present details of our approach and simulations including sections on i) the modeling of Q&A users including their associated traits and behaviors which we have extrapolated from our Yahoo! answers dataset[**?**]. ii) our single hop routing tactics which provide plausible deniability iii) an overview of our ant like trail laying/following stigmergic routing based approaches iv) high level details of simulations including key metrics finally section v) with a selection of graphs.

References

[1] Y. W. D. C. (L6), "Yahoo! answers comprehensive questions and answers version 1.0," 2009.

Collaborative Privacy Policy Authoring in a Social Networking Context

Ryan Wishart (Imperial College London)

Recent years have seen a significant increase in the popularity of social networking services. These online services en- able users to construct groups of contacts, referred to as friends, with which they can share digital content and communicate. This sharing is actively encouraged by the social networking services, with users' privacy often seen as a secondary concern. In this paper we first propose a privacy-aware social networking service and then introduce a collaborative approach to authoring privacy policies for the service. In addressing user privacy, our approach takes into account the needs of all parties affected by the disclosure of information and digital content.

Distance Matters: Geo-social Metrics for Online Social Networks

Salvatore Scellato

Abstract

In the recent years massive Online Social Networks (OSNs) such as Facebook, LinkedIn and Twitter have become increasingly popular, gathering millions of users and engaging them in the production, sharing and consumption of information over social links. These OSNs are increasingly becoming location-aware: they offer the opportunity to share geographic location in order to generate location-tagged information and to search for it. These new features open novel research directions which are largely unexplored, such as the design of new social applications and the improvement of existing large scale systems. Hence, it becomes important to investigate how geographic distance between individuals affects OSNs in order to deepen our understanding of these networks.

In this work we present a new approach for the analysis of networks with geographic information: we define new geo-social metrics which are able to quantify the geographic properties of the social ties across people. We describe a framework where network nodes are embedded in a metric space, in order to study the relationship between social connections and geographic distance. We define two new geo-social measures: a *node locality* metric, which quantifies how much a node is engaged with a local rather than global set of individuals, and a *geographic clustering coefficient*, which extends the standard notion of clustering by taking into account how much clusters of people are connected by short-range ties.

We apply our metrics to four different OSNs which provide location information for their users: we study two purely location-based social networking services (BrightKite and FourSquare), one blogging community (LiveJournal) and a social micro-blogging platform (Twitter). In particular, our results show how the new generation of location-based services services is characterized by short-range friendship links among users, resulting in a vast proportion of them with high values of node locality. On the contrary, these patterns are not present in social networks which are less centered on user location: in LiveJournal users have connections with heterogeneous length and this effect is even greater in Twitter. Morever, we find that location-based OSNs such as FourSquare and BrightKite tend to have more geographically confined triangles than social networks more focused on content production and sharing such as LiveJournal and Twitter, while standard clustering coefficient is not affected by this distinction.

These results indicate how taking into account geographic distance in these metrics provides insightful information for the design of systems and applications that could potentially exploit the underlying geographic and social structure of OSNs.

Empowering people to fight killer flu with mobile phones and social networking

Chris Watkins
Department of Computer Science, Royal Holloway.

Introduction

This is a discussion poster on possible ways to use mobile communications and Internet-based social networking to empower people to avoid getting infected during an epidemic.

In the 19th Century it was the civil engineers who did the most to banish disease, by delivering clean water for drinking and washing, and by building drains to separate foul water from the clean.

In the 21st Century there is a different challenge: in a world of crowded cities and rapid global travel, how to defend against a sudden global epidemic of a new disease? Once again, it may be engineers who provide the best defense, this time using applications for mobile phones and social networking.

New infectious diseases emerge frequently, but most of them fail to spread. Epidemics are either huge or tiny, and most are tiny. But there is a risk that a new infectious disease could emerge and cause a sudden, severe epidemic, spreading rapidly all over the world. When SARS emerged in 2003, doctors were prompt and heroic in tracing contacts and containing the epidemic before it spread. Will we be so lucky next time? What can be done to prepare?

It would be worthwhile to develop software tools that people could use to protect themselves, and that local communities could use to coordinate.

Communications technology could be used in (at least) four ways during an epidemic: for **reduction of social contacts** (`social distancing'); for **real-time epidemiology**; for **estimating personal risk of infection**; and **in enabling local communities to communicate and coordinate** mutual assistance, without having to hold public meetings to set such a system up. Ideally, software applications for all four of these uses should be developed on a precautionary basis, before any epidemic starts.

Reduction of Social Contact: 'Social Distancing'

To prevent an epidemic from spreading, it is sufficient to reduce the average number of people that each sick person infects. This number varies over time during an epidemic: at time t it is denoted R_t At the start of an epidemic, $R_0 > 1$, but as the epidemic continues, R_t falls because the fraction of susceptible, uninfected people falls. When R_t falls below 1, the epidemic peters out.

There are numerous ways in which people could use technology to reduce social contact, and hence reduce R_t : every small reduction would help.

During an epidemic, social distancing is common sense. Nevertheless it may be useful to prepare some software applications in advance of any epidemic, so that people will not have to invent and develop them after an epidemic has started.

For example, it would be worthwhile to develop an application that would enable any small shop to put its stock on-line, and to accept customer orders, as quickly and simply as possible.

Real-Time Epidemiology

Data from social networks and from the localisation of mobile phones could in principle provide very rapid information about the recent contacts of newly reported or self-diagnosed cases of the disease. From the point of view of epidemiology, this information would be of unprecedented detail and coverage, and might enable modes of disease transmission to be identified in detail.

Techniques for gathering this information, protocols for preserving confidentiality personal information, and methods of analysing the information, need to be developed beforehand.

Estimation of Personal Risk of Infection

Social distancing is costly and inconvenient, and people will not be able to reduce social contact effectively for long. It would therefore be desirable for people to be able to estimate their personal risk of infection from day to day, based on accurate real-time data of local prevalence of disease, so that people make an informed decision as to what precautions to take.

Local data from real-time epidemiology would in principle be a detailed resource that people could use to estimate their own risk of infection: once again, issues concerning confidentiality and rules for data handling, and methods of summarising and displaying it in a form that people could use, need to be developed.

Coordination of Local Community Response

During an epidemic, local communities would need to coordinate ways that were not previously necessary. Communities – such as the residents of an apartment complex or of a street – would need to:

- coordinate care for the sick
- support families in household quarantine
- coordinate to avoid mutual infection (eg in use of lifts in a tower block)
- respond to law-breaking and other emergencies if normal emergency services are over-stretched

In may city streets or apartment complexes, the residents may not even know each other: in this case, they would need to initiate cooperation without having large public meetings.

It would be common sense to use social networking software to initiate cooperation – but this software has not been designed with this application in mind. It may be worthwhile to consider the requirements of community cooperation during an epidemic, and to ensure that social networking

software does have the necessary privacy features to enable to people to join local groups and organise them, without having to give away

Conclusion

During an epidemic, social distancing, real-time contact tracing, estimates of personal risk, and methods for community organisation would be complementary and would reinforce each other in mitigating the epidemic. Software support for these activities should be provided in advance.

If an epidemic of a new, highly contagious disease should break out, people will spontaneously attempt to avoid being infected by making use of mobile and internet communications – this would be common sense. But people could be empowered to protect themselves better if suitable applications were developed and tested in preparation before any epidemic started.

Growth mechanisms in continuously-observed networks: Communication in a Facebook-like community

Tore Opsahl (Imperial College Business School, Imperial College London)

Most network studies rely on static data, which creates methodological issues when predicting tie creation. Although there has been a surge in continuously-observed datasets (i.e., non-static), few methods exist to study these. This paper proposes a framework for assessing multiple growth mechanisms (e.g., homophily, focus constraints, reinforcement, reciprocity, triadic closure, and preferential attachment) in such datasets, and applies it to communication within a Facebook-like community. While some mechanisms are significant independently, they are insignificant in multivariate analyses. This finding exemplifies that descriptive measures, such as the clustering coefficient, cannot be relied upon for studying mechanisms of tie generation.

See arXiv:1010.2141 (http://arxiv.org/abs/1010.2141) for paper.

How social media helps the music industry

Tracking artist popularity, fan interactions and media consumption using Musicmetric

Trung Huynh, Gregory Mead, Jameel Syed, Matthew Jeffery.
Musicmetric - www.musicmetric.com

Introduction

Over the past decade, social networks have been rapidly becoming the most popular platform for people to communicate and discuss their favourite artists, songs and concerts. The online footprint left by music fans is getting larger by the day, and by tracking this information in real time we can supply the music industry with valuable data on artist popularity, consumer behaviour, fan interactions and opinions. A few examples of how this information can help the music industry include: targeted marketing, optimising efficiency of marketing campaigns, discovery of emerging artists, identifying and targeting 'super fans' and minimising damage from piracy.

In this poster we explore some of the methods used by Musicmetric to track online behaviour of music consumers on social networks and the web, and how this information is used by the music industry.

Musicmetric is a technology start-up based in London, UK that tracks all aspects of music consumption online including: social network monitoring, web crawling and sentiment analysis, peer to peer download tracking, identifying influential fans, supplying fan demographic information and geographic data on where an artist is popular. Our customers include record labels, marketing companies, advertising agencies and broadcasters in the music industry. The sections of the full poster paper are outlined below, and will cover in more detail the methods of analysis used in each part.

Fan interaction on social networks

Musicmetric tracks the number of plays, views and comments on artists' MySpace profiles, number of fans on Facebook, number of related tweets on Twitter, number of plays on last.fm and several other social popular networks. The poster will show the types of time-series analysis performed on this data.

Fan profile on social networks

Musicmetric analyses information about fans on social networks and produces an overview about artists' fan demographic and geographic information. The poster will cover how to efficiently collect this data.

Fan influence on social networks

Musicmetric is also able to tell which fans of an artist are most influential in terms of how frequent they interact on social networks, how many other influential friends they have and their Eigen centrality.

Fan opinions on social networks

Musicmetric analyses reviews, blogs, comments, tweets of professionals and fans on social networks and the whole world-wide-web to produce an overview of what Internet users think about artists' latest tracks, albums, concerts, etc. The algorithms and benchmarks for results are covered in detail on the poster.

Other areas

Musicmetric also tracks peer-to-peer downloads by region, online radio stations, and music related web sites. These will be covered in less detail as they are not directly relevant to social network analysis.

Imperial College
London
BUSINESS SCHOOL

The proposed poster intends to provide an overview of the research on the role of "Networks in Open Innovation", undertaken at Imperial College Business School. A summary of this line of research can be found below:

Open Innovation: how corporate R&D scientists capture and exploit external knowledge from their networks

Many organizations have attempted to open up their innovation processes by increasingly exploiting external sources of knowledge to help them generate, develop and commercialize innovative ideas. These search activities range from creating and maintaining external connections with universities, suppliers, and customers to attending conferences and technology fairs with the aim of identifying new technologies and products and to promote their development within the firm. These open innovation initiatives require a new set of skills and practices, and many firms have struggled to train and manage their employees to become more externally facing. The ability of scientists, engineers and technical staff to routinely reach beyond the boundaries of the firm for new ideas and to integrate them inside the firm has become a critical organizational challenge. Despite the central role of external engagement in innovation, we still know very little about successful search and networking practices associated with open innovation. Moreover, we lack the tools to assess the ability of individuals to recognize the value of external ideas and to integrate them inside the organization.

Our research aims to improve our understanding of how corporate R&D scientists develop new ideas and subsequently transform them into successful products. R&D scientists are increasingly expected to function as 'gatekeepers', proactively gathering external information and then translating this information in terms that are meaningful and useful to other members of the organization. To be able to perform this function gatekeepers need to be strongly connected both to other members of their organization as well as having rich contacts to external sources of information.

This project identifies the search and networking practices of scientists and engineers in a large multinational company by mapping their internal and external networks and measuring successful instances of integration of externally sourced ideas. The core of the research project is a large-scale survey of senior scientists and engineers in the technical career ladder of this company. The survey has been preceded by a series of interviews with senior R&D scientists.

A better insight in the role of networks in Open Innovation allows formulating lessons on how R&D scientists can more effectively utilize their networks to create breakthrough innovations. Further, the results of this study generate greater understanding on how firms can enhance the search and networking activities of their R&D staff and design effective strategies for sourcing and integrating external ideas, a critical component of a firm's approach to open innovation.

Dr. Anne ter Wal - Dr. Paola Criscuolo - Prof. Ammon Salter

Innovation Studies Centre
Imperial College Business School

Privacy concerns and social network routing

Iain Parris and Tristan Henderson
School of Computer Science
University of St Andrews
St Andrews, Fife, KY16 9SX, UK

November 10, 2010

Abstract

Researchers have proposed novel Internet applications for social network information beyond "traditional" social network sites, e.g., search, distributed computation, and security. Users' privacy concerns are typically overlooked when considering such novel applications.

We present a case study of the performance of social network routing in opportunistic networks, before and after considering users' privacy preferences. We find that taking users' privacy concerns into account may lead to dramatically lower performance for social network routing schemes.

Poster

We would be interested in presenting the following poster during the SOCIALNETS workshop on 18 November: `http://www.cs.st-andrews.ac.uk/~ip/tmp/2010-11-18_socialnets_parris.pdf`

We previously presented this poster during the informal SCONE meeting on 14 September, and are keen to receive further feedback from workshop attendees.

Thank you for your consideration.

Contact details

Iain Parris: ip@cs.st-andrews.ac.uk

Small-world behavior in time-varying graphs[1]

John Tang

Abstract

The analysis of social and technological networks has attracted a lot of attention as social networking applications and mobile sensing devices have given us a wealth of real data. Classic studies looked at analysing static or aggregated networks, i.e., networks that do not change over time or built as the results of aggregation of information over a certain period of time. Given the soaring collections of measurements related to very large, real network traces, researchers are quickly starting to realise that connections are inherently varying over time and exhibit more dimensionality than static analysis can capture.

In particular, the small-world phenomenon, i.e., the fact that real networks have high clustering coefficient, while the typical distance between their nodes is small as in random graphs, has been investigated in *static graphs*, neglecting the temporal dimension [2, 3, 6]. The time evolution of a real system, when considered, is usually studied by evaluating the standard static measures (distances and clustering coefficient) on snapshots of the network taken at different times [5, 4]. However, this approach does not capture entirely the dynamic correlations of a time-varying network.

In this poster we will present firstly, new temporal distance metrics to quantify and compare the speed (delay) of information diffusion processes taking into account the evolution of a network from a global view and secondly, a measure of how fast a dynamic network evolves. Intuition would lead us to believe that slowly evolving networks would be slower for information diffusion, however on the contrary using both a modified brownian motion model and empirical traces we find that time-varying networks can be strongly clustered in time and, at the same time, exhibit short temporal paths between their nodes. This phenomena has important dynamical consequences for biological, social and man-made systems and we hope that our work will stimulate further studies of temporal small-world behavior in real time-varying systems.

References

[1] J. Tang, S. Scellato, M. Musolesi, C. Mascolo, V. Latora, Physical Review E, Vol. 81 (5), 055101, May 2010.

[2] D. J. Watts and S. H. Strogatz, Nature **393**, 440 (1998).

[3] V. Latora and M. Marchiori, Phys. Rev. Lett. **87**, 198701 (2001).

[4] A. Clauset and N. Eagle. Persistence and Periodicity in a Dynamic Proximity Network. In *Proc. of DIMACS Workshop on Comp. Methods for Dyn. Int. Nets.*, (2007).

[5] A.-L. Barabási, H. Jeong, R. Ravasz, Z. Neda, T. Vicsek, and A. Schubert, Physica A **311**, 590 (2002).

[6] C. Caretta Cartozo and P. De Los Rios Phys. Rev. Lett. **102**, 238703 (2009).

Studying Location Sharing on Social Networks with Mobile Experience Sampling

Fehmi Ben Abdesslem, Iain Parris and Tristan Henderson
School of Computer Science, University of St Andrews, UK
{fehmi,ip,tristan}@cs.st-andrews.ac.uk

1. INTRODUCTION

Formal interviews and questionnaires allow to collect self-reported information about users' behaviours when using mobile social applications, but users may forget some details about their experiences or report inaccurate information when answering questionnaires. On the other hand, analysing the information shared on social network sites (SNSes) only allows the examination of those information that have been shared, rather than the information that have not been shared, or the contexts in which users do not wish to share. The Experience Sampling Method (ESM) [3] is a diary method that consists of asking participants to stop at certain times and report about their experiences in real time.

We advocate the use of ESM, possibly in addition to questionnaires and analyses of SNS accounts, for capturing information about mobile users' behaviour *in situ*, when the mobile social application is actually used. In this poster, we describe our testbed using ESM with smartphones to study the behaviour of mobile social application users.

2. EXPERIMENT

Our research is interested in how, when, where and to whom people share their locations with their social network, to better understand their privacy concerns. We go a step further than previous experiments by actually disclosing location to the participants' social network. Moreover, we use a single device to detect location, ask ESM questions, and then collect both ESM answers and detected locations.

Our experiment [1] involved 80 participants sharing their location to their social network with a smartphone over the course of one week. Each participant was given a Nokia N95 8GB smartphone, constantly running a custom application that detects their location using GPS and Wi-Fi scanning. Locations were regularly uploaded to our server through the cellular network, and published on their Facebook SNS account according to their disclosure choices. To this end, participants were asked during a prebriefing session to set up friend groups on Facebook if these did not already exist (e.g., family, classmates) and default disclosure choices. Signal-contingent ESM questions (at random times of the day) and event-contingent ones (when the server detects a new location) were sent to the participants through an SMS handled and displayed by the application (cf. Figure 1).

3. CHALLENGES

Compared to SNS analysis or traditional surveys, implementing the Experience Sampling Method to study the behaviour of mobile social application users is more complicated and time consuming. Our method requires designing, implementing and deploying an appropriate testbed composed of smartphones to collect data and a server to monitor and store these data.

Figure 1: A participant is asked whether he would share a photograph with his social network friends.

Using a single device to collect data, ask questions and collect answers necessitates the use of more energy than the normal use of such a device to answer calls. In particular, monitoring users' behaviour continuously may involve multiple sensors to be triggered frequently, which may quickly deplete the battery. Hence, managing efficiently the sensors to save energy is an important challenge to collect data on participants' behaviour in their everyday lives. For instance, in our system, we use the accelerometer embedded in most smartphones to detect motion, and switch off the GPS when the participant is not moving to save energy [2].

Other challenges include avoiding the experiment being too intrusive, remotely managing the devices and ethical considerations that must be carefully taken into account.

Acknowledgements

This work is funded by the EPSRC/TSB *Privacy Value Networks* project (EP/G002606/1).

4. REFERENCES

[1] F. Ben Abdesslem, I. Parris, and T. Henderson. Mobile experience sampling: Reaching the parts of Facebook other methods cannot reach. In *Proceedings of PUMP*, Dundee, UK, Sept. 2010.

[2] F. Ben Abdesslem, A. Phillips, and T. Henderson. Less is more: energy-efficient mobile sensing with SenseLess. In *Proceedings of ACM MobiHeld*, pages 61–62, Barcelona, Spain, Aug. 2009.

[3] R. Larson and M. Csikszentmihalyi. The experience sampling method. *New Directions for Methodology of Social and Behavioral Science*, 15:41–56, 1983.

EXTENDED ABSTRACT

Jisun An, Computer Laboratory, University of Cambridge

Recently online social networking sites like Facebook and Twitter have emerged as a popular way of discovering information on the World Wide Web. In contrast to traditional methods of content discovery, such as browsing or searching, content sharing in social networking sites occurs through word-of-mouth, i.e., content spreads by the way of conversations between users. For instance, users share links (URLs) to content on the Web with remarks like "Have you seen this article?" or "Check out this YouTube video." This new way of discovering information has rapidly gained popularity and today social networking sites are known to be a major driver of traffic to many web sites. Recent data from the Twitter network shows that nearly 23 million web links are shared every day.

To better understand this newly popular phenomenon, in this paper, we present a detailed analysis of word-of-mouth exchange of URLs between Twitter users. With over several tens of millions of tweets per day and over a quarter of them containing URLs, Twitter provides a particularly rich data set to study word-of-mouth based URL discovery. Our analysis investigates and characterizes several aspects of word-of-mouth based content discovery including its impact on URL popularity and its effect on the diversity of information discovered by users.

In order to study how the 54 million users in Twitter collaboratively discovered and spread web links, we built an information propagation tree for every URL that was shared during a random week in 2009. We used Krackhardt's hierarchical tree model to construct the paths of information flow.

Based on the propagation trees of URLs, we answer a number of questions that are fundamental to understanding word-of-mouth based web discovery. The questions we ask include: Can word-of-mouth reach a wide audience? What kinds content are popular in social media as opposed to the web in general? Does word-of-mouth give all content, including those published by unpopular domains, a chance to spread? What are the typical structures of word-of-mouth propagation trees? We discuss the implications of our findings for the design of word-of-mouth based marketing strategies and the role of social media.