Wired to Connect: Analyzing Human Communication and Information Sharing Behavior during Extreme Events

Faisal Aleissa
Center for Complex Engineering Systems at KACST and MIT
fsaleissa@kacst.edu.sa

Riyadh Alnasser
Center for Complex Engineering Systems at KACST and MIT
ralnasser@kacst.edu.sa

Abdullah Almaatouq
Center for Complex Engineering Systems at KACST and MIT
amaatouq@mit.edu

Kamran Jamshaid
Center for Complex Engineering Systems at KACST and MIT
kamjam@acm.org

Fahad Alhasoun
Center for Computational Engineering, Massachusetts Institute of Technology
fha@mit.edu

Marta C. González
Department of Civil and Environmental Engineering, MIT, Cambridge, USA
martag@mit.edu

Anas Alfaris
Center for Complex Engineering Systems at KACST and MIT
anas@mit.edu

ABSTRACT

In this work, we study the communication dynamics and information propagation of real-world events through the contact networks of mobile phone users. Previous studies have shown that these 'bread-crumbs' of digital traces can act as in situ sensors for human behavior, allowing for quantifying social actions and conducting social studies at an unprecedented scale. However, most work in utilizing these proxies has focused on the study of human dynamics under regular and stationary situations, with little research on the quantitative understanding of human behavior under extreme events. In this work, we examine three events with different size and geographical scope and show that (i) human communications are both temporally and spatially localized during such events; and (ii) various types of events produce a distinct human communications signature both on a temporal and a spatial scale.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Communications Applications, Miscellaneous

Keywords

Social Norms, Activity patterns, Call Detail Records

1. INTRODUCTION

Humans have an innate need to connect and share information with others. This is especially profound during times of rare events, when we are anxious to communicate with our loved ones. Studying the dynamics of human communications and information sharing behavior in these circumstances has multiple practical benefits; we are particularly interested in its application to enhance the capacity of public safety organizations. We motivate this as follows:

Imagine a scenario where some set of individuals witness an extreme event (e.g., heavy rains, a large fire outbreak, etc.). These individuals then communicate about this event with others in their social circles. At the same time, various law enforcement and public safety organizations are at work trying to quickly assess the scope and size of the event. Early on, when the situation is volatile and unclear, these agencies may be able to benefit from tools that can aggregate and analyze the intensity and dynamics of 'chatter' across various communication mediums generated by witnesses on the ground. In this paper we address the question whether it is possible for external observers (such as public safety agencies) to witness the dynamics of this communication, without necessarily monitoring the actual content, and still be able to make an inference about the scope and the size of the event?

A sizable portion of the world's population currently utilizes mobile phones and social media (e.g., Twitter, Facebook) for much of their communications [1, 8]. Previous work [5, 7, 10, 11] have shown that these 'bread-crumbs' of digital traces can act as in situ sensors for human behavior; allowing for quantifying social actions and conducting social studies on an unprecedented scale. However, most of this work in utilizing such proxies has focused on studying the human dynamics under regular and stationary situations, with little work on quantitative understanding of human behavior under extreme events [6, 12].

In this work, we study the human dynamics and information propagation of real-world events through the contact networks of mobile phone users. We utilize a large dataset of Call Detail Records (CDRs), which are kept by telecommunication companies for billing purposes, as a proxy into human behavior during emergencies. Every time a user makes a phone call, sends a text message, or uses the Internet, the mobile network keeps a record of their usage information and location [2], with all personal data being anonymized.

We start by extracting three different extreme events from our CDR dataset. The first is a fire outbreak at a remote waste treatment plant near an industrial city in a Middle Eastern country, resulting in several fatalities and injuries. The second event is a major sports fixture with more than 25,000 attendees. The third event is a weather scare event where the forecast of heavy rains prompted early school closures, creating considerable consternation and inconvenience among concerned parents. These events can broadly be categorized into two categories based on prior awareness about the event as well as its scope and size: (i) events whose existence was known in advance to a large number of people (both the sports fixture and the rain scare event in a major metropolitan city); and (ii) events whose existence was unforeseen and was limited in their effect (the fire outbreak in a remote factory).

Our results in this paper show that these types of events trigger an anomalous change in behavior and a spike in the communication activity of those who witnessed it. Although this phenomena has been observed in previous studies (i.e., [3, 4, 6]), we show that the three events have a unique signature as reflected in their spatio-temporal analysis. This signature reflects both the size and the scope of the event in terms of number of people who communicated about it and can be used to distinguish events.

Our main contributions in this paper are as follow: using spatio-temporal analysis of the human communication signature pattern across different types of extreme events, we show that:

- Human communications are both temporally and spatially localized during such events.
- Various events produce a distinct human communications signature both on a temporal and a spatial scale.

2. DATA DESCRIPTION

The dataset consists of one full month of records for an entire country, with 3 billion mobile activities to over 10 thousands unique cell towers, provided by a single telecom service provider. Each record contains an anonymized user ID, the type of activity (i.e., SMS, MMS, call, data etc), the cell tower ID facilitating the service, duration, and time stamp of the activity. Each cell tower ID is spatially mapped to its latitude and longitude. For privacy considerations, user ID information has been anonymized by the telecom operator.

Previous studies [7, 9] have shown that human communication patterns are highly heterogeneous, with some users using their mobile phone much more frequently than others. The characteristics of the dynamics of individual communication activity obtained in Fig. 1 supports such hypothesis. Fig 1a shows the Empirical Cumulative Distribution Function (ECDF) of the activities duration. We find that almost 75% of the duration of the conducted activities last for 70 seconds or less. Fig. 1b shows the statistical distribution of the number of communication records generated by the users for a single day. Fig 1c shows the inter-event time distribution $P(\Delta t)$ of calling activity, where Δt is the time elapsed between consecutive communication records (outgoing phone calls and SMS) for the same user. These results conform to the findings of previous studies [7, 9], which increases our confidence in the validity of our dataset.

3. PRELIMINARIES

In this section, we introduce the terminology that we use in our work:

- We first identify the start and end time (t_{start} and t_{end}, respectively) for the three events. For the factory explosion, both t_{start} and t_{end} correspond to the time of the explosion. For the football match, these correspond to the match start and end times, respectively. For our weather event, we used the historical hourly weather data to identify t_{start} and t_{end} for the rain in the first half of the day.
- We define event duration as a time interval spanning 30 minutes before t_{start} and 90 minutes past t_{end}.
- Witnesses are users in the vicinity of the event who communicated during the event duration, as registered in our dataset. While we have no means to verify the content of their communication, considering their proximity and the significance of the event, we assume that the witnesses chose to communicate about the event.
- For our baseline reference, we use activities preceding 3 working days before the event. This was necessitated by several factors: (i) We use working days since the activity levels on weekends are considerably different; (ii) we use preceding working days since some types of events alter user behavior after the event. For example, following the factory explosion, the closest cell towers show a reduced activity level for the remainder of the month. We suspect this may be because some workers were asked to stay at home or report to a different site while the factory repairs were being performed.

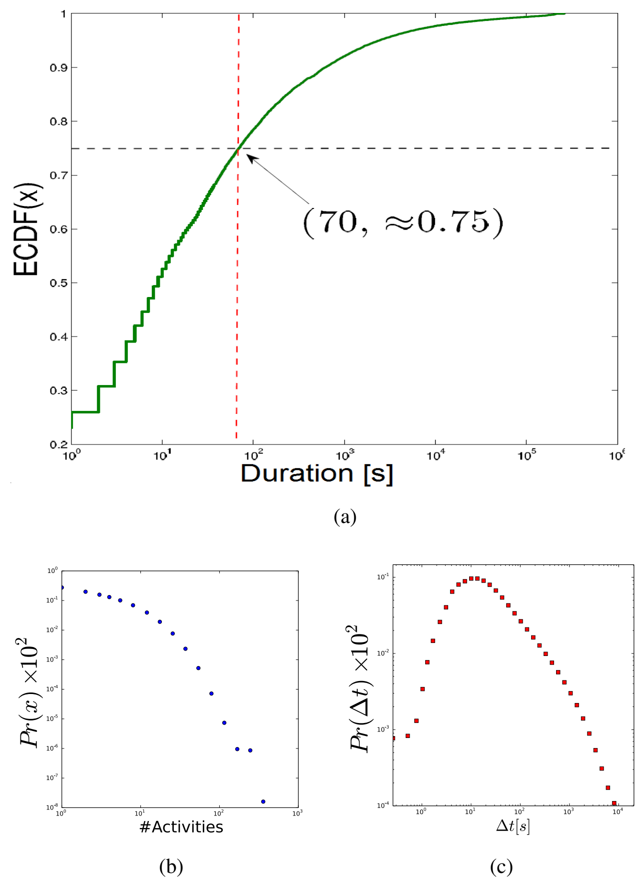

Figure 1: Communication patterns in the CDRs dataset

4. TEMPORAL BEHAVIORAL CHANGE

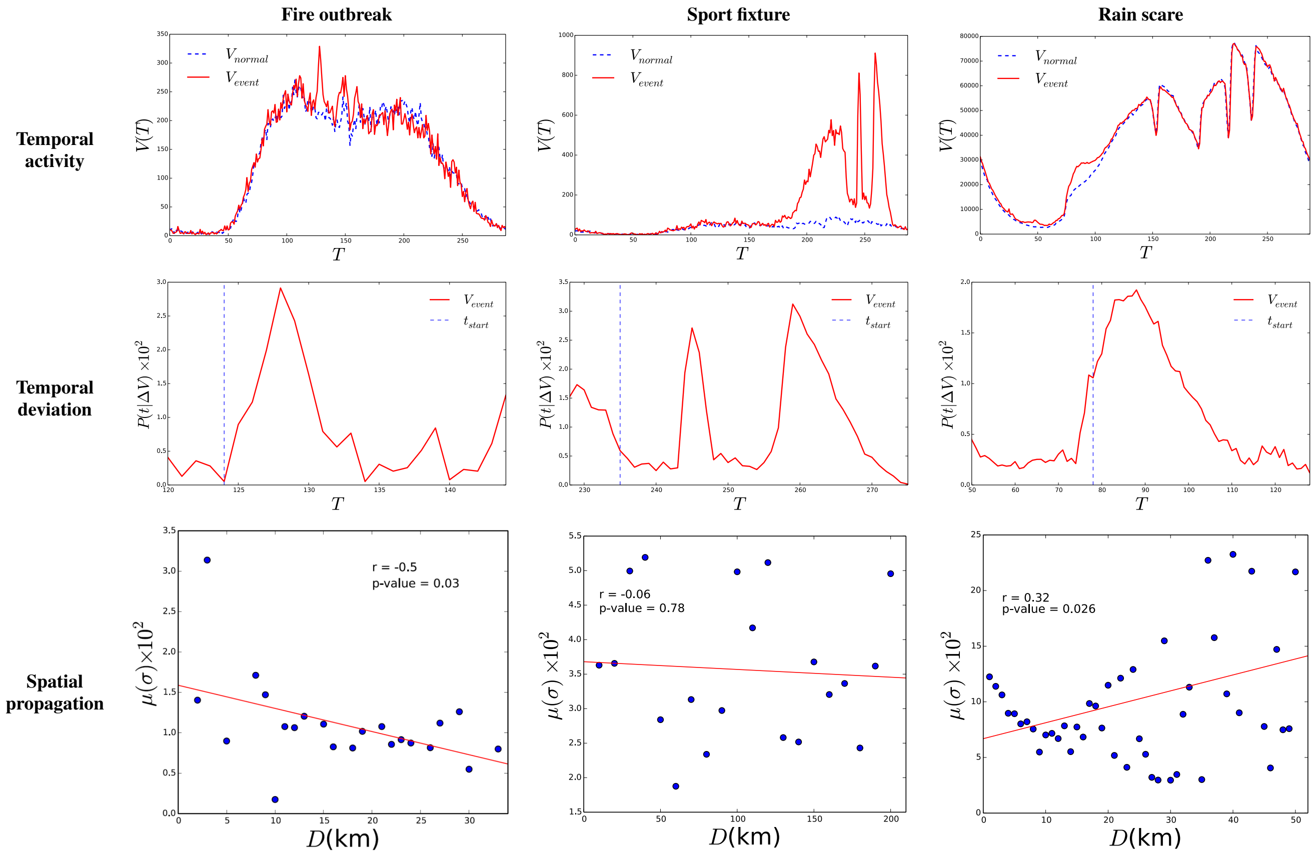

Figure 2: The influence of a fire outbreak, sport fixture, and rain scare on the temporal and spatial patterns of mobile phone activity

In order to analyze the impact of extreme events on communication patterns, we log the activities observed by the cell towers closest to location of incidents. We decompose the daily mobile phone activities into 288 five minutes interval. We quantify the magnitude of the effect of the incident on the communications activity by reporting the variation of the activity level from the norm at these cell towers. The normal pattern is determined by averaging the activity across three preceding business days, as described in Section 3. We denote the normalized activity levels during the day of the event and normal days as V_{event} and V_{normal}, respectively. Our results in Fig 2 show that there exists an abnormal behavior in calling activity where the nature of the event seems to control both the temporal and spatial propagation of the event.

The first row in Fig 2 shows the temporal behavior of calling volume for the emergency event V_{event} and normal days V_{normal}. The second row shows the deviation from the mean for the events observed. The deviation is quantified as the following:

$$P(t|\Delta V) = \frac{|V_{event} - V_{normal}|}{\sum V_{normal}}$$

where $\Delta V = |V_{event} - V_{normal}|$, the vertical dashed line in the figure represents the start time of the event t_{start}. Fixing the spatial dimension during the analysis highlights the localized temporal behavior of the mobile phone activity levels.

Fig 2 shows signatures of the three events on the temporal and spatial dimensions. First, a fire outbreak shows a spike in phone call activity occurring right after the incident. The plot shows that the variation in activity from the normal daily activities is significantly large. We also observe that the change in activity pattern is temporally localized, with high activity levels for around 30 minutes before returning to normal levels.

For the sports fixture, we observe three distinct spikes in communication activities. The first spike corresponds to the gradual build-up with the arrival of spectators prior to t_{start}; the second deviation corresponds to the peak at half time (roughly 45 minutes past t_{start}), while the third spike corresponds to the end of the match.

For the rain scare event, we observe that the communication patterns also reflect a build-up before t_{start}, the rain start time as obtained from historical weather records. This is because of the weather forecast, many schools started calling the parents before the rains started to alert them that they are cancelling classes for the day. News reports suggest that this happened earlier in the morning, at some schools within a short time of when the parents had just dropped off their children, resulting in traffic congestion and considerably inconvenience.

5. SPATIAL BEHAVIORAL PROPAGATION

In order to study the spatial propagation of the event, for each event we observed the deviation in activity between the cell towers in close proximity of the event to the surrounding cell towers. We used the Haversine formula to calculate the shortest distance D of a cell tower from the event (ignoring any hills, etc.)

After observing an event and finding its closest cell tower, we define the set of towers belonging to a certain radius d as follows:

$$\sigma_d = \{c_1, c_2, c_3, ..., c_n \ : \ |dist(c_{event}, c_i) - d| < \alpha\}$$

where $dist(c_{event}, c_n)$ is the Haversine distance between the cell tower closest to an event c_{event} and any other cell towers denoted as c_i. Hence, σ_d is the set of cell towers that are of distance between $d - \alpha$ and $d + \alpha$ from the closest tower to an event c_{event}. We then quantify the average deviation around a distance d with a threshold α as the following:

$$\mu(\sigma_d) = \frac{\sum_{c \in \sigma_d} P(t|\Delta V)}{|\sigma_d|}$$

The spatial propagation of the events, as measured, is shown in the last row of Fig. 2. The correlation coefficient r and the p-value for each event is also listed. Similar to the temporal behavioral change, we observe that the spatial behavioral propagation for the three events also shows a distinct signature. For the fire outbreak in a remote factory, the plot shows a negative linear correlation between the variation in calling activity and the distance from the incident. Events such as fire outbreaks where immediate attention is required are very highly localized spatially and temporally. Incidentally, the highest change in activity is observed at a tower about $3km$ from the site of the event, but close to the housing facilities of the workers at the plant. We suspect that this reflects calls made by families and friends to inquire about the well-being of their loved ones.

The sports fixture shows weak spatial localization ($r = -0.06$ and $r > 0.05$). We suspect that this is because popular sports fixtures with large fan followings are broadcast live on TV, and thus the impact of such fixtures reflects a wider spatial propagation courtesy of the 'virtual witnesses', all of whom may not be at the actual scene of the event.

While it is easy to identify the geographical point-of-activity for the fire outbreak and the sports event, such identification may not be possible for events that impact the entire city, such as severe weather. For the rain scare event, we show the spatial propagation of the event from the city center to its periphery in steps of $1km$. Unlike the previous two events, we observe a positive correlation coefficient with increasing distance; this is because many schools wishing to expand their facilities are choosing to relocate to new development areas at the periphery of the city rather than the city center to save on their costs.

For the three events, we observe that the fire outbreak is most narrowly localized in space (similar to what has been shown by [4], in a boom attack event), unlike football matches that are not especially localized or city scale events such as weather alerts.

6. SUMMARY AND FUTURE DIRECTION

A sound quantitative understanding of human behavior under duress is essential for a number of practical problems faced by emergency responders, ranging from emergency detection to traffic control and management. In this work we showed that various types of events produce a distinct human communication signatures both on a temporal and a spatial scale.

This analysis using our CDR dataset may reflect some of the properties specific to our dataset. For example, the mobile phone penetration rate for our Middle Eastern country is up to 180%. Thus, it is not uncommon for a person to own two or more mobile phone numbers. Second, our CDR data is spatially mapped to the locations of the cell towers, and not the actual locations of the users making the phone calls, and therefore the coordinates of cell towers are used as a proxy to the locations of users. Finally, since we have only a one month worth of data, we limit our comparative baseline to use data for 3 working days prior to the event as some events occur early during the month.

In our on-going work, we are building quantitative understanding of the dynamics of links creation and network evolution during extreme events. In addition, we will investigate the universality of behavioral changes. For example, does it hold across types of events, cultures and demographics? We will also attempt to understand the dynamics of the social network and on the information network during extreme events.

7. REFERENCES

[1] A. Almaatouq, A. Alabdulkareem, M. Nouh, E. Shmueli, M. Alsaleh, V. K. Singh, A. Alarifi, A. Alfaris, and A. S. Pentland. Twitter: Who gets caught? observed trends in social micro-blogging spam. In *Proceedings of the 2014 ACM Conference on Web Science*, WebSci '14, pages 33–41, New York, NY, USA, 2014. ACM.

[2] A. Almaatouq, F. Alhasoun, R. Campari, and A. Alfaris. The influence of social norms on synchronous versus asynchronous communication technologies. In *Proceedings of the 1st ACM International Workshop on Personal Data Meets Distributed Multimedia*, PDM '13, pages 39–42, New York, NY, USA, 2013. ACM.

[3] Y. Altshuler, M. Fire, E. Shmueli, Y. Elovici, A. Bruckstein, A. Pentland, and D. Lazer. The social amplifier–reaction of human communities to emergencies. *Journal of Statistical Physics*, 152(3):399–418, 2013.

[4] J. Bagrow, D. Wang, and A. Barabasi. Collective response of human populations to large-scale emergencies. *PloS one*, 6(3):1–8, 2011.

[5] N. Eagle and A. (Sandy) Pentland. Reality mining: sensing complex social systems. *Personal Ubiquitous Comput.*, 10(4):255–268, Mar. 2006.

[6] L. Gao, C. Song, Z. Gao, A.-L. Barabási, J. P. Bagrow, and D. Wang. Quantifying information flow during emergencies. *Scientific reports*, 4:3997, Jan. 2014.

[7] M. C. Gonzalez, C. A. Hidalgo, and A.-L. Barabasi. Understanding individual human mobility patterns. *Nature*, 453(7196):779–782, June 2008.

[8] M. Nouh, A. Almaatouq, A. Alabdulkareem, V. Singh, E. Shmueli, M. Alsaleh, A. Alarifi, A. Alfaris, and A. Pentland. Social information leakage: Effects of awareness and peer pressure on user behavior. In *Human Aspects of Information Security, Privacy, and Trust*, volume 8533 of *Lecture Notes in Computer Science*, pages 352–360. Springer International Publishing, 2014.

[9] J. Onnela, J. Saramaki, J. Hyvonen, G. Szabo, D. Lazer, K. Kaski, J. Kertesz, and A. L. Barabasi. Structure and tie strengths in mobile communication networks. *Proc. Natl. Acad. Sci. USA*, 104(18):7332–7336, 2007.

[10] A. Pentland. *Social Physics: How Good Ideas Spread - Lessons From A New Science*. Peguin Press, 2014.

[11] M. Vieira, E. Frias-Martinez, P. Bakalov, V. Frias-Martinez, and V. Tsotras. Querying spatio-temporal patterns in mobile phone-call databases. In *Mobile Data Management, 2010 Eleventh International Conference on*, pages 239–248, 2010.

[12] O. Woolley-Meza, D. Grady, C. Thiemann, J. Bagrow, and D. Brockmann. EyjafjallajÃűkull and 9/11: The impact of large-scale disasters on worldwide mobility. *PloS one*, 8(8), 2013.