
Community Level Health Data: Looking at Real-Time Dating Sites

Mark Handel

(The Boeing Company)
Seattle, WA
handel@gmail.com

Irina Shklovski

IT University of Copenhagen
Copenhagen, Denmark
irsh@itu.dk

Abstract

We discuss the potential for a new source of data for personal health maintenance, community-level data sourced from a variety of social networking sites. We present some preliminary results on extracting this kind of data from dating sites catering to men who have sex with men (MSM), and discuss possible user communities for this information.

Author Keywords

Community Health Sensors; MSM; HIV; Real-Time Dating Sites

ACM Classification Keywords

H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

Introduction

Most of the research in the quantified-self movement has focused on individual level data, which can be used to improve health and well-being. These data points have the advantage that they are acquired and reported by the individual who is the direct user of the relevant devices, and there are only "classical" privacy concerns within the data ecosystem: who has access, how are the data summarized and aggregated, and so forth. However, self-reported health information and quantification of self does not necessarily have to be

limited to the use of particular devices (such as FitBit for example) or mobility tracking services (such as Strava). As sharing of health information becomes more normatively acceptable, it is likely we will begin seeing certain types of information disclosed on regular social network sites or dating sites. In fact, this is already happening in certain niche communities, such as among men who have sex with men (MSM) where particular types of health information, such as HIV status – can be vitally important in the context of interaction and especially dating. However, data obtained from such sources as online dating profiles comes with a different issue – dating profiles tend to be exceedingly stable and thus the health data can be significantly out of date [1]. Yet we argue that these kinds of data, nevertheless, can provide a relatively accurate measure of community-level data. We discuss some of our preliminary work in a novel area, but one that has fairly important health implications, at least for the members of the specific community under study.

We have been studying real-time dating sites (RTDS) [1], including, but not limited to location-based systems (e.g. Grindr and Scruff) and websites (e.g. Manhunt and Gaydar). These online dating sites are primarily geared towards MSM, and although there are a range of sites, including mobile apps, most of them have a core set of features: individual profiles, an “online now” indicator, geographic restrictions, as well as a messaging infrastructure. Within the MSM community, some of these sites and apps are very popular (we will refer to them generically as sites): the biggest of the mobile apps have over 1M active users, and the web sites can also boast a similar reach. With such a wide user base, there are obviously differences in goals and usage patterns; some people use them for

casual sexual encounters, while others seek more of a community orientation even when using products that support new social interactions [2].

Dating Sites as Data Sources

The dating technologies popular in the MSM community are interesting from a well-being standpoint because they act as a new source for certain types of health information. Although this information is generally about an individual user, we want to focus on the potential for using this data as a community-level indicator or sensor for important health-related information. This is an area of emerging interest, for instance, recent work has explored the potential of Craigslist for providing sensors for sexual-health related content in personal ads [4]. These sites gather a wide range of both self-reported and implicit data about their users. The self-reported data often includes the expected questions about age, weight, height as well as drinking and smoking but also items more specific to this community, such as personal HIV status, potential partner HIV status and drug use information. Obviously, self-reported data has significant limitations and biases, but many of these biases are relatively well understood, for instance, weight is often under-reported especially on dating websites.

In addition to these self-reported data, there are also large amounts of implicit data that are generated by individuals’ use of these products. These encompass such things as temporal patterns of usage, short or long-term absences from product usage, as well as overall intensity of usage. These can be collected both on an individual level as well as aggregated across varying levels of community. In addition, there is the potential to examine various kinds of communities:

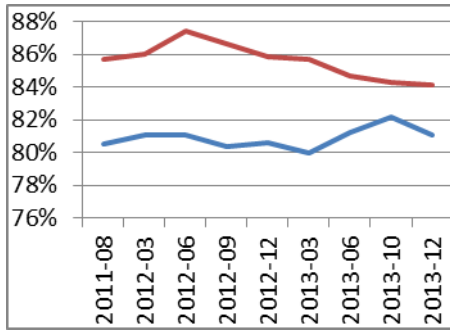


Figure 1: HIV Prevalence in two cities over time

geographic distribution is an obvious contender, since this is how many of the products are structured, but there is also the potential to cluster on other relevant factors, such as age, HIV status or drug use.

Through our work we have found that both classes of data are, within the terms of use restrictions, relatively easy to gather and process. We have used the data gathered to look at the potential for creating these sensors. We will look at some potential designs, and then discuss how these sensors could be deployed to a variety of stakeholders: both individuals and organizations.

Our Work

We have been examining data collected from these sites to look at three potential community-level signals: HIV status, drug use, as well as overall usage intensity. For the MSM community, HIV status is obviously a very important indicator for health and well-being. However, HIV status is not necessarily the dichotomous variable it once was. Modern HIV treatment regimens can bring the viral load down so low as to be undetectable, which greatly reduces the risk of transmission. More recently an increasing number of HIV- individuals have begun taking a Pre-exposure Prophylaxis regimen (PrEP), which can also greatly reduce the risk of transmission. Thus, just saying positive or negative is no longer necessarily enough to fully capture the range of risk. In addition, research has shown that individuals who are not aware of their HIV status are often the biggest source for new HIV infections [5]. This has led to a new push for regular testing among sexually active MSM. Although most of the sites under study have controlled vocabulary fields and drop-down menus to indicate HIV status at varying levels of granularity almost none have a formal "Recent Test Date (RTD)" field. Nevertheless,

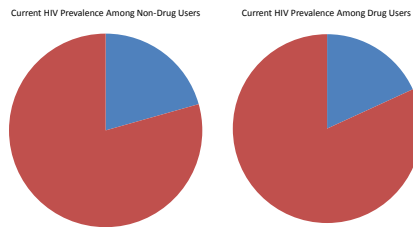


Figure 2: HIV Prevalence, clustered by drug use

we have been observing and increasing number of users putting dates of their last test in the free text sections of their profiles.

In addition to HIV status, another major impact to overall health is the rate of drug use. Although most sites list this as "drug use," in practice, this is usually understood to mean methamphetamine use, a popular drug in the MSM community (It is not unusual for a profile to say no drugs, but also indicate openness to "softer" drugs like marijuana.) Like HIV status, drug use status is often a mixture of controlled vocabulary fields in the profile, mixed in with additional details in free-text profiles.

We have been looking at ways to aggregate these disparate pieces of information, which can be missing, incomplete, or inaccurate from the profiles, and using it to give an overall metric of certain health aspects of a community. Much as vaccination can provide for herd immunity, many of the HIV and drug use harm reduction efforts can also provide for increased overall risk reduction.

We are still exploring the design space, in terms of aggregation and presentation of this data, as well as exploring how users will react to this kind of information. In figure 1, we show a graph of HIV prevalence over time in two different cities. Another possible visualization is to cluster results along other factors, such as drug use, as seen in Figure 2. (Unfortunately, in this instance, there is only a 2% difference between the two populations)

Discussion and Future Steps

These sensors can be used by individuals to get a better sense of the community they are interacting with. This information can be used to help them moderate their own actions, or to provide prompts for discussion before initiating intimate contact. In addition, they can also be used by community health advocates to understand areas for outreach and interventions. For instance, if the average time since the last HIV test is growing, additional testing opportunities can be scheduled.

One obvious area of concern is anonymity in this data. In our examples, we have been careful to only present data at a community level; most of the results are the combination of hundreds of individuals over many sampling points. However, it would be interesting to discuss these kinds of sensors with users of the sites. An interesting tension that may arise is that, individually, all of this data comes from "public" data on the site: this is information that is present in their profiles, and visible to all of the other users of the site. However, the aggregation may be the source of concern, even though it is less identifiable than the individual profile data.

Bibliography

[1] N. Ellison, J. Hancock and C. Toma, "Profile as Promise: A Framework for Conceptualizing Veracity

in Online Dating Self-Presentations," *New Media & Society*, vol. 14, no. 1, pp. 45-62, 2012.

- [2] M. Handel and I. Shklovski, "Disclosure, Ambiguity and risk reduction in real-time dating sites," in *Proceedings of the 17th Annual ACM International Conference on Supporting Group Work*, Sanibel Island, FL, 2012.
- [3] M. Handel, J. Birnholtz and I. Shklovski, "Grindr-ing Down the Walls Between Physical and Virtual: Location-Based Social Enablers," in *iConference 2014*, Berlin, 2014.
- [4] O. Haimson, J. Brubaker and G. Hayes, "DDF Seeks Same: Sexual Health-Related Language in Online Personal Ads for Men who have Sex With Men," in *Human Factors in Computing Systems (CHI)*, Toronto, 2014.
- [5] S. Cassels, T. W. Menza, S. M. Goodreau and M. R. Golden, "HIV serosorting as a harm reduction strategy: evidence from Seattle, Washington," *AIDS*, vol. 23, no. 18, pp. 2497-2506, 2009.