

Rethinking Kelly and Etling's Map of the Iranian Blogosphere
Gwyneth Sutherland, Doctoral Researcher
University of Bradford, UK
Using Open Data: Policy Modelling, Citizen Empowerment, Data Journalism
Brussels, 19 May 2012

A few guiding questions

What are the limitations of open data for IR policy?

What is the macro framework which informs the approach to mining open data for social science?

What is a valid framework when working with global communications data?

How can culture be incorporated as a variable?

Introduction

Kelly and Etling of Harvard's Berkman Institute conducted a three-part series for the *Internet and Democracy* project mapping online environments in regions of strategic importance to American policy-makers which led to a further research initiative with the United States Institute for Peace called *Blogs and Bullets*. This paper takes a closer look at *Mapping Iran's Online Public* (2008) because it is a good exemplar of the prevailing methodology in both the series and in the field of open data use for policy making.

The original study analyzed open data in the form of blog URLs and associated links. In the first stage, the sites were visualized using a Fruchterman-Rheingold 'physics model' algorithm. The resulting clusters, called poles, were named and described through a text-mining filter designed by selecting 1700 terms "of interest" from en.wikipedia.org which also had an associated Farsi translation. (Kelly and Etling, 2008, p.15) Several native-level Farsi speakers reviewed hundreds of blogs by hand and coded for topics and information about authors. Finally, some associated links, such as YouTube videos, were considered as *outlink* analysis which looked at density of links connecting to other information sources to form a larger online ecology while still adhering to the network visualization model of nodes, poles, and links.

The foundation of the map, what Kelly and Etling call the 'macro structure' of their analyses hinges on social science research about American behavior toward information and social group formation which they assert can be extended to other cultures and to the activity of blogging. (2008, p.8) In both of these regards, they overreach. Following from the flawed macro structure, their methodology produces an invalid modeling of Iranian online politics. This paper proposes a critique and a few tentative suggestions which highlight the value of *culture* as a variable in analyzing communication data.

Universal Limitations

Kelly and Etling (2008, p.6) claim that:

Unique as a snowflake, the network structure of a society's blogosphere will reflect salient features of the society's culture, politics, and history. A society's online communities of interest, social factions, and major preoccupations can be seen and measured, their words read and analyzed through a combination of structural and statistical analysis and textual interpretation.

And they further assert that:

Understanding the map is the key to understanding the Iranian blogosphere. (p.7)

If this network is unique, then why underpin the analysis with a macro structure hybridized from two social science theories that assert all humans, regardless of cultural background, behave similarly? Put more simply, proposing that the system has universal and predictable qualities and is also a unique snowflake is a difficult model to build. Analysis of data filtered through an online platform, such as a blog, frequently ignores the invisible variable of cultural translation or

context. We have been primed by the idea of globalization. Anything we all use must be used in the same way. There is a sense of equivalency, of shared experience. Research begins with the false assumption that data transmitted through this platform have a universality because a perceived quality of the technology has been collapsed with that of the data transmitted across it.

The two theories Kelly and Etling used as the foundation to their 'physics model' map combined conclusions about communication bias of Americans in the late 1950s and 1960s who had selective exposure to information with a concept on how groups or networks (socially, not online) coalesce because of affinity also based on studies done in an American context.

1. Sociology has extensive literature on *homophily*, the tendency of social actors to form ties with similar others.
2. Communications research has identified complex processes of *selective exposure*, by which people chose what media to experience, interpret what is experienced, and remember or forget the experience according to their prior beliefs. (Kelly and Etling, 2008, p.8)

In both of these quotations there is a reliance on universal qualities of 'people' applied to a group of people we admittedly have a weak understanding about. The social science theory among social network theories called *homophily*, McPherson, Smith-Lovin, and Cook (2001) explains simply, "similarity breeds connection" in the introduction to their survey of research exploring group formation among the heterogeneous, often contentious, population of the United States. During decades of racial integration and political remapping, understanding what held American society together and caused groups to form, produced several network theories which have resurfaced to make sense of online communities. (Borgatti et al., 2004) However, these theories were not meant to explain sociopolitical dynamics in other cultures. Extending their analysis further to understand online sociopolitical behavior in other cultures is considerably beyond what these theories can support. Luna et al (2002) explored several applications of culture as a variable when restructuring website navigation flow and interface design done by business marketing researchers. Motivated by financial success, companies found online users behaved differently in ways that researchers aligned with cultural markers or values.

Iran has historically posed a challenge for Western intelligence gathering and policy making. Events there have frequently caught the outside world by surprise. If we can concede that the subject is unfamiliar, then forcing the data to conform to a familiar visualization tool or metaphor displays more loyalty to the model than to what can be learned from the wealth of new data available.

Visualizing a system as a network, as a series of linear links with nodes or poles limits the ways we can discuss relationships in that system. We become constrained by the metaphor which channels conclusions towards causal linkages and presents distinct antipodes dichotomizing the landscape. When seen as a global or spherical view it encourages thinking that this space shares a geography. The connections may in fact be across the diaspora, which is significant for political analysis (location of sites was added to the mapping of the Arab Blogosphere later in the series). Is this the true nature of the system or the image created by a visualization tool with insufficient means to describe that system? Certainly all models fall short in some respect, but for a poorly understood political landscape such as Iran, collapsing the data to fit a model which is understood in American contexts will not advance understanding of Iranian contexts. Above all, remember that the data represents communication; they are not a neutral things. Building a model which captures qualities of that cultural element, communication, will enrich our understanding of circumstances beyond imposing a one-size-fits-all concept of data.

Proposal

1. The 'attentive clusters' depicting the 'informational worlds' (p.6) place enormous importance on information gathered from online sources in Iran. Do Iranian's use blogs or the web to get information? How?

2. Rate of change. Rather than the initial visualization of the 'physics model,' begin with a semantic filter for unusual words, such as is done with abstract construction, and performed during known social/political events. Some events might be local which could help determine location of bloggers, some might be internationally covered, which could place bloggers within larger information ecology. Word change over a time period could indicate engagement role in information flow. However, these words would be selected for their uniqueness, or other non-political quality in order to let the political context of the bloggers emerge of its own accord. These terms may lead to more information about bloggers such as age and location based on their uniqueness.
3. How to make sense of cultural context? Kelly and Etling dismiss two elements worth pursuing as distinct markers for Iranian online discourse. First, how do strategies to avoid censorship affect online discourse? Anecdotally, much online text is not to be trusted, there is considerable nuance, subtlety and 'code' used to convey meaning in all forms of public discourse such as film and offline written materials. In fact, many of the clusters and results which did not conform to the polar/network model were discarded. If the data suggested the model, they may have contributed new understandings. Second, the orality of the Iranian discourse is not easily mapped to online written sites. Iran remains a place where information travels my word of mouth, then by phone call, and more complex still, the type of information and the age or socioeconomic status of the individual may determine still how the information travels. Weighting the relative importance of online discourse within the larger political discourse might be measured with traffic flow or mining text changes. There could be a study soon which maps where Farsi Wikipedia entries originate. This written format participation could be compared with YouTube contributions or traffic to understand written vs. oral communication preferences online.

Conclusion

Policy makers are concerned with creating a visualization tool or a model with the data in order to facilitate predictions. The possibility for this is extremely limited when the architecture of the model rests on social science. The model remains, no matter how many variables contribute, one of an incredibly complex set of interactions between human beings who do always not respect rational outcomes. What a good model can offer is a rich description of the current environment without the promise of predicting how factors affect or might manipulate that environment. We do not yet have enough of these descriptions of online spaces. They would serve policy-makers whose judgment and ability to assimilate information outweighs our current capacity to build models.

References

- Borgatti, S., Mehra, A., Brass, D. and Labianca, G. (2009) Network Analysis in the Social Sciences. *Science* [online] pp. 892-895. Available at: DOI:10.1126/science.1165821 [Accessed 12 May, 2012].
- Kelly, J. Etling, B. (2008) Mapping Iran's Online Public: Politics and Culture in the Persian Blogosphere. Berkman Center Research Publication [online]. Available at: http://cyber.law.harvard.edu/publications/2008/Mapping_Irans_Online_Public [Accessed 27 July 2011].
- Luna, D., Peracchio, L., and de Juan, M., (2002) Cross-Cultural and Cognitive Web site Navigation. *Journal of the Academy of Marketing Science* [online], vol. 30(4), pp. 397-410. Available at: 10.1177/009207003236913 [Accessed 7 November 2011].
- McPherson, M., Smith-Lovin, L and Cook, J. (2001) Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology* [online], vol. 27, pp. 415-444. Available at:

<http://www.jstor.org/stable/2678628> [Accessed 13 May 2012].

Sears, D. and Freedman, J. (1967) Selective Exposure to Information: A Critical Review
The Public Opinion Quarterly [online] vol. 31(2), pp. 194-213. Available at:
<http://www.jstor.org/stable/2747198> [Accessed 10 May 2012].