# Quantification of Topic Propagation Using Percolation Theory: A Study of the ICWSM Network

**Ali Azimi Bolourian** and **Yashar Moshfeghi** and **C. J. van Rijsbergen**
Department of Computing Science, University of Glasgow
Sir Alwyn Williams Building, 18 Lilybank Gardens
Glasgow G12 8QQ, Scotland, Great Britain
{azimi, yashar, keith}@dcs.gla.ac.uk

**Abstract**

Blogs facilitate online debates and discussions for millions of people around the world. Identifying the most popular and prevailing topics discussed in the Blogosphere is a crucial task. This paper describes our novel approach to the quantification of the level of topic propagation in the Blogosphere. It tries to answer one key question: *How many people should know about a subject before it becomes prevalent?* Our model uses graph-theoretic representations of the Blogosphere's link structures that allows it to deduce the 'Percolation Threshold', which is then used in the quantification and definition of a prevalent or 'Global' topic. We applied our approach and analysed the social structure of the ICWSM data collection to find the answer.

## 1. Introduction

Ubiquity and affordability of access to the Internet together with Web 2.0 technologies have contributed to a surge in the popularity of online social networks and creation of user-generated content on the web including blogs. Despite this massive growth (Kelleher 2008), the industry has not been able to effectively harness the power of social networks due to the fact that role and effect of social media are not well studied and understood (Kelleher 2008). What is required are models that study and model the social dimension in such networks.

Several models and approaches have been proposed trying to model various aspects of the blog network. These models cover a broad range of features in the Blogosphere, from analysing dynamics and trends (Llora et al. ), trust and influence (Kale 2007) to community evolution (Lin et al. 2007) and genre analysis (Herring et al. 2004). Research work such as (Adar and Adamic 2005) has looked into tracking information epidemic in the Blogosphere. The increasing number of research publication in this field emphasises the importance of the need to understand the behaviour and dynamics of the Blogosphere. Despite the various existing approaches that analyse the diffusion of information, there has not been any clear definition on the metrics of propagation.

Our research is the first attempt to give an accurate measure for the level of information propagation. This paper presents 'SugarCube', a model designed to tackle part of this problem by offering a mathematically precise solution for the quantification of the level of topic propagation. The paper also covers the application of SugarCube in the analysis of the social network structure of the ICWSM/Spinn3r dataset (ICWSM 2009). It presents threshold values for the communities found within the collection, and paves the way for the measurement of topic propagation within those communities. Not only can SugarCube quantify the proliferation level of topics, but it also helps to identify 'heavily-propagated' or *Global* topics. This novel approach is inspired by Percolation Theory and its application in Physics (Efros 1986).

The remainder of this paper is structured as follows. Section 2 describes our motivation in pursuing research on measurement of topic propagation in the Blogosphere. Section 3 gives a brief introduction to Percolation Theory and follows by discussing topic percolation in the context of blogs. In Section 4 we introduce the model and methodology of SugarCube, and define the notion of a 'global topic'. In Section 5 we describe our experiment and findings with the ICWSM dataset. We present our findings in Section 6, conclude in Section 7, and briefly review future work ideas in Section 8.



## 2. Motivation and Application

The social network space, including social media and social networking web sites have heavily influenced the way the Internet is used, and has affected culture, business, politics, and virtually every aspect of modern life. Yet the social networking phenomenon has also created new challenges for both industry and academic research. For businesses, one social networking metric is distinctly underwhelming: companies have not been able to effectively harness the power of social networks for revenue generation. All the attraction and interaction in the social networks did not lead to a dramatic increase in the earning of classical online advertising models. Despite the ongoing research and apparent business potentials, many social networking sites are having trouble capitalising on their audience. Even a company like Google that has excelled in the business of online advertisement, is stopping to pay a premium rate for ads displayed on social networking sites such as MySpace. The industry has started to look into new models of online advertisement where ads are not only relevant to the content of their placeholder, but also have a social dimension (Kaplan 2008) (Ante, Grover, and Green 2007).

Moreover, the substantial increase in the content of social media is putting pressure on information retrieval systems and search engines. In a quest to provide the best possible service, search engines attempt to index everything. The ever increasing volume of social content requires not only a shift in indexing policies, but also a change in the way hardware, software, and human resources are assigned to different tasks. A search engine that can measure and predict the prevalence of subjects would be able to allocate human and computing resources more effectively to indexing, processing and storage of data and various other tasks. In the online advertisement business, for instance, an ad campaign with limited budget would be able to make optimal use of their assets and resources by knowing the number of people they need to be reaching before their campaign is considered to be a success in their target market. Likewise, an online service would be able to offer improved personalisation services when the social network of its client base and their interests are well-defined.

To sum up, despite this massive growth in social media, the industry has not been able to effectively harness the power of social networks due to the fact that role and effect of social media are not well studied and understood. (Kaplan 2008) (Ante, Grover, and Green 2007) (Kelleher 2008). What is required are models that study the social dimension in such networks. This research aims to address part of these problems, i.e. help add a new social dimension by quantifying propagation levels. It provides the answer to this key question: *How many people should know about a subject before it becomes prevalent?*

## 3. Percolation Theory

Since the inception of Percolation Theory (PT) by Broadbent and Hammersley (Frisch and Hammersley 1963) in 1957, the theory has been used to interpret an exceptionally wide variety of physical and chemical phenomena. From formation of the electrical properties of amorphous semiconductors and spontaneous magnetisation of a doped ferromagnetic, to spread epidemics and forest fires, Percolation Theory has been widely used in several branches of science (Efros 1986).

The problems best tackled by Percolation Theory are *critical phenomena*, which are characterised by a *critical point* at which some of the properties of the system undergo abrupt changes, such as the transition of a metal from its normal form to its superconducting phase, or the spread of a manageable forest fire to an uncontrollable fire engulfing the forest (Efros 1986) (Stauffer 1994).

Despite the variation and irregularities in the physics of critical phenomena, all of them share a common feature: in the neighbourhood of the critical point, the system breaks into blocks with different properties. The size of the blocks grows until the system reaches the critical point. There is no regularity in the geometry of these blocks in the vicinity of the critical point. However, due to the large size of the blocks, the geometry is independent of the structure of the system. Therefore systems of varied nature and structure may behave similarly when they approach the critical point. It can therefore be said that the "geometry of disorder determines a number of properties of a system in the vicinity of its critical point." (Efros 1986)

Percolation Theory takes advantage of this relationship between physics and geometry. Different structures and geometries can be associated to a certain lattice structure that allows a precise application of the statistical-physical technique of PT. A number of lattices are well studied, and their PT-related features are precisely known (Jeraldo 2008).

There are two basic types of percolation models. In the first one, points are defined on an underlying lattice in such a way that, in every lattice site there is a probability $p$ for a point to exist there. This is the site percolation model. In the second one, bonds are defined between two neighboring sites on a lattice. Each bond has a probability $p$ to exist. Accordingly, this model is called bond percolation. In both cases, structures of connected points can be defined, in a way that it is possible to create a path between any two points of the cluster.

The value of the critical probability $P_c$ in which the system would go through abrupt changes can be calculated precisely for some lattices in both site and bond percolation. For the rest, $P_c$ has no precise answer and numerical and computational methods are used to derive useful approximates (Essam 1980) (Kesten 1982) (Stauffer 1994).

### Topic Percolation in the Context of Blogs

The Blogosphere is not just a place to publish and express opinions. Blogs facilitate online debates and discussions for millions of people around the world. Some blog posts go unnoticed and some attract the attention of other bloggers and readers, and fuel debates and discussions.

The aim of this work is to find out the number of people who should know about a subject before it becomes prevalent, i.e. the *critical* fraction of bloggers who are "actively aware" of the topic. Active awareness is the situation in which bloggers not only know about a certain subject, but also engage in online debates and discussions about it and use hyperlinks during the course of their activity. In other words, active awareness is reflected in the Blogosphere by the act of linking. Based on this definition, blog entries with no in-coming and out-going links are ignored. To achieve this, we first look into the subjects which are actively discussed in the Blogosphere, count the number of bloggers engaged in each topic, and find out if it has attracted a 'critical mass'.

In this context, it is useful to look at the network of blogs as a graph, in which the vertices are blogs (representing bloggers) and the edges are friendship (blogroll) links connecting vertices. Figure 1 illustrates such a graph along with different levels of topic propagation: some topics emerge in a corner of the Blogosphere and spread to only a few other bloggers (Zone A in Figure 1); some attract more attention, reach a larger group of actively-aware audience but eventually cease to propagate further and fail to capture the attention of a wider public (Zone B in Figure 1); and finally, those topics that spread all over the Blogosphere (Zone C in Figure 1). We call this type of highly propagated topic a global topic. Our objective is to define what level of spread constitutes a global topic, and develop a model based on PT to quantify it.

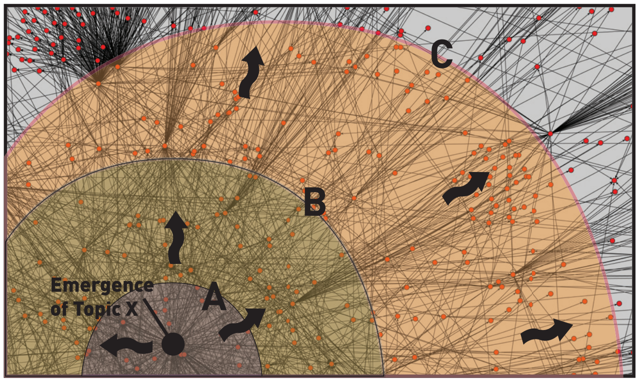


Figure 1: Schematic view of the propagation of topics within a large network of blogs: Some topics only attract a few bloggers (A), some reach a larger audience but fail to capture public attention (B), and some reach a 'critical mass' and become 'global' (C)

This is not the first use of Percolation Theory in modelling social trends. Ahmed and Abdusalam (Ahmed and Abdusalam 2000), Takahashi and Murai (Takahashi 2004) and Efros (Efros 1986) have proposed models based on Percolation Theory for various social phenomena, from the spread of rumours in a local community to the social relationship in the context of politics. However, to the best of our knowledge this is the first time a model is proposed for the analysis of social structure in the Blogosphere based on Percolation Theory.

In the following sections, the notion of global topic will be defined and expanded within the framework of SugarCube model.

## 4. Methodology and Model

SugarCube focuses on hyperlink analysis and differentiates between friendship links and blog entry links as the approach in (Leskovec and McGlohon 2007). It utilises the combination of the following three relationship structures represented in graph-theoretic form:

1. the graph of *Blogroll Links* (BL) also known as friendship or readership links;
2. the graph of blog *Entry Links* (EL);
3. a Percolation Graph (PG).

In SugarCube, BL and EL graphs are two hyperlink layers that together constitute the Blogosphere, as shown in Figure 2. These three graphs are described in more detail below.

### Blogroll Link Graph (BL)

Blogroll Link graph, also known as friendship or readership graph, is a graph $G_{BL}(V, E)$ where vertices are blogs (representing bloggers), and edges are blogroll or friendship links between blogs. $G_{BL}$ represents the social network structure of bloggers, provides valuable insight into each blogger's position within the blogging community, and reveals their likely sources of influence and information. The spread of information within the social network depends on the connectedness amongst bloggers which is reflected in the structure of the BL graph.

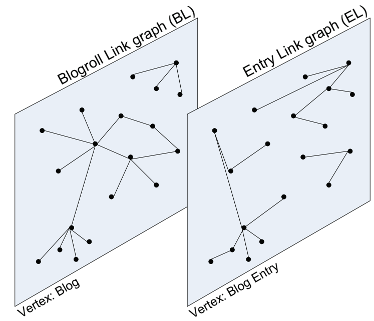


Figure 2: The Blogosphere consists of two hyperlink layers: the Blogroll Links (BL) and the Entry Links (EL)

Evidently, bloggers might have ways other than blogrolls to access news and information (such as links they keep in their browser bookmarks or receive in emails). However, blogroll links are very explicit signs of bloggers' preferred connections and possible sources of influence.

The structure of the BL graph plays a significant role in the spread of topics, and consequently is an important factor in determining the critical threshold in the SC model. The probability for information to move from one person to another within a well-connected community is much higher than in a hypothetical community where people do not know each other and never talk to one another. Therefore, the potential for topics to spread throughout a community such as the Blogosphere depends on the level of connectivity between its members, namely the bloggers. This connectivity is explicitly reflected in the BL graph, and will be used in configuring the Percolation Graph later in this paper.

### Entry Link Graph (EL)

EL is a graph $G_{EL}(V, E)$ comprising of blog entries (represented by permalinks). EL edges are usually formed by means of direct linking, commenting and trackbacks. As illustrated in Figure 3, the EL graph is represented on a 2-dimensional plane called Blog-Time in which X-axis is Time ($T$), Y-axis consists of Blogs ($B$) in an arbitrary order, and each permalink is a vertex $v(B, T)$ where $v \in V$ in the graph $G_{EL}$, denoting the blog and time of the entry. Figure 3 schematically shows a number of weakly connected components (WCC) in Blog-Time. SugarCube intuitively assumes that all the vertices in a particular subgraph are discussing the same topic: that's why they linked to each other.
The EL graph answers three key questions:

- How many weakly-connected components or topic-based subgraphs exist?
- How many blogs are involved in each subgraph?

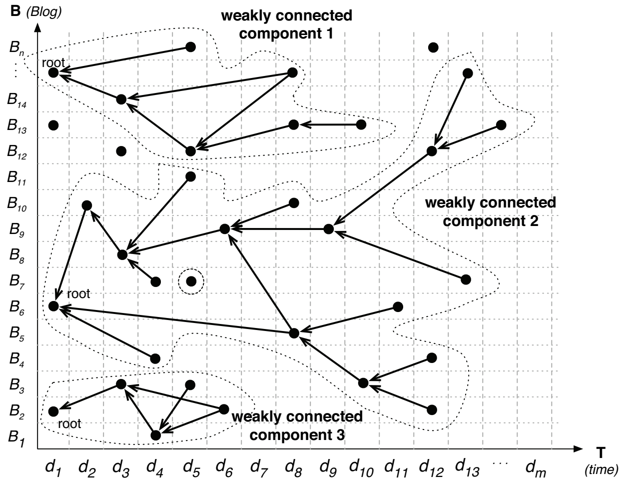


Figure 3: Weakly connected components formed around topics in the EL graph.

- What is the time-span of each discussion?

For instance, three weakly-connected components 1, 2 and 3 can be seen in Figure 3. Even though the discussion in $WCC_3$ consists of five entries, the number of blogs involved in this discussion is only three: all five entries are posted by blogs $B_1$, $B_2$, and $B_3$. It is also evident that the discussion in $WCC_3$ went on for six days. Similarly, the discussion in $WCC_2$ consists of 18 entries posted by 13 blogs during a time span of 14 days. Consequently, the topic of $WCC_2$ has reached twice as many blogs as the topic of $WCC_3$. In other words, the topic of $WCC_2$ has attracted more attention and percolated farther through the community of bloggers than the other topics.

The Percolation Graph introduced in the next section provides a threshold or an index used in the identification of heavily propagated or global topics.

## Percolation Graph (PG)

Inspired by the Rumour Propagation model of A. L. Efros in (Efros 1986), we adopt a lattice-based Percolation graph in order to transform the propagation problem into a Percolation Theory problem. The lattice chosen for this purpose is a Bethe lattice. This lattice was introduced by Hans Albrecht Bethe in 1935 within the context of statistical mechanics, and since then it has been used as an approximation for models in 2 and higher dimensions (Bethe 1935) (Braga, Sanchis, and Schieber 2005).

The Bethe lattice was chosen for three main reasons: percolation in the Bethe lattice starts from its "root" outwards, resembling the spread of information from one person to the wider social network. Also, the Bethe lattice has a configurable structure, its co-ordination number $z$ (number of immediate neighbours of a lattice vertex) can be set according to requirements. Moreover, due to its distinctive topological structure, the statistical properties of the Bethe lattice are exactly solvable, and computationally inexpensive. A simple Bethe lattice is shown in Figure 4.

In SugarCube, the value of $z$ in PG is approximated to the average in-degree value of $G_{BL}$ which indicates the average number of readership for a blogger:

$$z \approx \frac{1}{|V|} \sum_{v \in V} \rho^{-}(v) \quad (1)$$

where $\rho^{-}(v)$ denotes the in-degree of vertex $v$. Vertices in the lattice represent blogs. Propagation of topics is designated as a bond percolation problem in which bonds symbolise the spread of active awareness from one person to acquaintances within $G_{BL}$ with probability $p$.

## Global Topic

The exact solution to the bond percolation problem on the Bethe lattice was first given by Fisher and Essam (1961) and subsequently by (Sykes and Essam 1963). Percolation Threshold $P_c$ is the critical probability, in the vicinity of which the topic percolates throughout the Blogosphere. In other words, a critical number of people are actively aware of the subject. $P_c$ for the Bethe lattice is precisely computable:

$$P_c = \frac{1}{z-1} \quad (2)$$

This indicates that if the probability $p$ of bond connectivity between vertices (bloggers) exceeds $P_c$, then information will percolate throughout the entire lattice. $P_c$ denotes a tipping point for the propagation of a topic, and when $p > P_c$ that topic is said to have heavily percolated through the community of bloggers, and thus is considered to be a "global topic". In SugarCube, $p_i$ is the fraction of blogs (or bloggers) in $G_{BL}$ associated with a weakly connected component $WCC_i$ in $G_{EL}$. Therefore, a topic is considered global in a community if the fraction of bloggers who are actively aware of it reaches $P_c$.

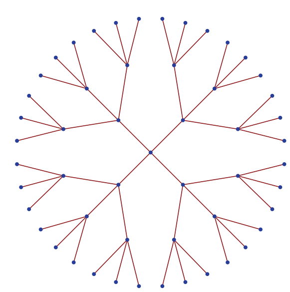

Figure 4: Bethe Lattice with $z = 4$

The next section describes our experiment with the ICWSM/Spinn3r collection.

# 5. Experiment Findings

## Step I - Identification of Blog URLs

We traversed the whole collection and created a set of home page URLs associated with all blog entry documents. A total of 2,627,194 unique blog homepage URLs were identified at this stage. Each blog homepage is assumed to represent a blogger (person).

Table 1: Basic statistics of the Blogroll graph $G_{BL}$

| Vertices | Non-singleton vertices | Edges |
|---|---|---|
| 2,627,194 | 1,089,223 | 11,839,763 |

### Step II - Finding Blogroll Links

Each and every homepage URLs was crawled to retrieve its set of blogroll (friendship) links. Out of all blogroll links found for each URL, only the subset of links that intersected the ICWSM collection were kept. This was to confine the boundaries of the experiment within the limits of the dataset.

We experienced a 1.7% failure rate during our crawls, i.e. could not retrieve home page data for 46,418 blogs. These include blogs that no longer exist, and servers that rejected our crawler bot.

With the exception of LiveJournal blogs, it was assumed that all Blogroll links reside on blog homepages. LiveJournal however provides a different mechanism for its members to express friendship links within the LiveJournal community. Around 30% (precisely 778,628) of the blogs in the collection are hosted by LiveJournal, thus we made sure those friendship links were extracted properly.

### Step III - Blogroll Graph

A Blogroll Graph $G_{BL}(V, E)$ was created where vertices are blogs (representing bloggers). Table 1 lists brief statistics on the number of vertices and edges of the graph. With 50,085 in-links, *http://jsp.typepad.com* has the largest in-degree in the collection.

Figure 5 illustrates the distribution of in-degrees in $G_{BL}$ on log-log scale. The head of the distribution follows a power law $P(k) \sim k^{-\alpha}$ with the exponent $\alpha = 1.03$.

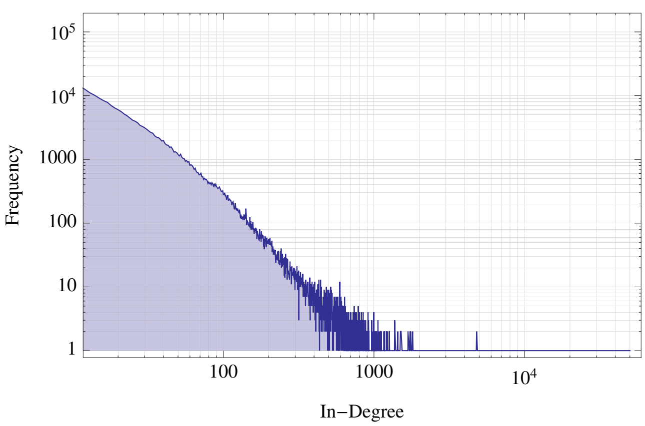


Figure 5: In-degree distribution of the BL graph

### Step IV - Communities

We identified 16,206 weakly-connected components (WCC) in $G_{BL}$, of which the largest one consists of 1,046,063 vertices. Table 2 lists the features of the top ten weakly connected components within the BL graph, together with the average in-degree of the component, chosen $z$ value for PG and the value of Percolation Threshold. Each component represents a community of bloggers in the collection. For each community presented in Table 2, the URL with the highest degree is chosen to identify the community. We also manually checked the identifier URL to see whether it *looks like* a real blog or splog, etc.

### Step V - Percolation Threshold

Based on the average in-degree value of each community, we assigned a coordination number $z$ to configure a suitable Percolation graph. Consequently, the percolation threshold $P_c$ for each community was calculated, which are listed in Table 2.

## 6. Discussion

Despite the relatively slow pace of change in the structure of the BL graph, changes in a small number of vertices (such as deletion of one highly-connected vertex) can dramatically alter the structural characteristics of the graph, which in turn, affects the configuration of the PG and consequently the value of $P_c$.

The goal of this paper is to provide a means for deducing the critical fraction of blogs that need to be overtaken for percolation to materialise. This in turn provides us with an index, to which we can compare the *level of percolation* or *level of globalness* of sub-global topics. If $n$ and $P_c$ are the existing and the critical fractions of actively aware bloggers respectively, then the scalar value $g$ could be defined as the level of globalness:

$$g = \frac{n}{P_c} \quad (3)$$

## 7. Conclusion

We proposed a new model for the measurement of topic propagation on the Blogosphere using Percolation Theory. We then put our theory to practice by analysing the social network within the ICWSM/Spinn3r collection. We then computed the value of Percolation Threshold for all communities. The thresholds show the fraction of people within each community who should know about a subject before the topic of that subject becomes prevalent, thereby providing an accurate answer to our research challenge. We proposed to use the notion 'global' for topics that pass the threshold to indicate their prevalence. We also showed a method for measuring the level of propagation of other (sub-global) topics by comparing them to the percolation threshold of the underlying community as an index.

To summarise, by undertaking this experiment we have:

- identified the potential application of Percolation Theory in the measurement of topic propagation in the Blogosphere;
- proposed a novel approach based on Percolation Theory that identifies the highly propagated topics, and defined the notion of a Global topic;
- applied our theory to the large-scale ICWSM/Spinn3r blog collection, extracted communities within the collec-

Table 2: Characteristics of the top ten communities within the ICWSM network of bloggers

| Community size | Average $\rho^-$ | $z$ | $P_c$ | Indentifier URL | Identified as |
|---|---|---|---|---|---|
| 1,046,063 | 11.2630 | 11 | 0.1000 | http://jsp.typepad.com | blog |
| 1,166 | 3.3490 | 3 | 0.5000 | http://financeausedcar.wordpress.com | splog |
| 117 | 8.5213 | 9 | 0.1250 | http://sellandtellonline.com | unknown[1] |
| 83 | 13.1445 | 13 | 0.0833 | http://megalizardfan.livejournal.com | blog |
| 72 | 22.6666 | 23 | 0.0454 | http://cbkopf.exblog.jp | splog |
| 64 | 2.6093 | 3 | 0.5000 | http://ownlink228.cocolog-nifty.com/blog | splog |
| 60 | 23.6 | 24 | 0.0434 | http://beaconbadge.livejournal.com | blog |
| 56 | 5.75 | 6 | 0.2000 | http://brummitt89.livejournal.com | blog |
| 50 | 44.22 | 44 | 0.0232 | http://tm675345.spaces.live.com | splog |
| 49 | 6.2653 | 6 | 0.2000 | http://debbie-g.livejournal.com | blog |

tion, computed Percolation Threshold for each community, and presented the results for the top 10 communities;

- paved the way for other researchers to further analyse the structure of the EL graph and draw conclusions on the propagation level of topics in the collection.

To the best of our knowledge, this is the first application of Percolation Theory in the Blogosphere.

### Further Work

**Analysis of the EL Graph** The next stage would be to analyse the EL graph by extracting the WCCs formed by active discussions, counting the number of bloggers involved in each discussion, and finding out if any of the discussions have attracted a 'critical mass'. We are currently undertaking further work into the ICWSM/Spinn3r collection to analyse the EL graph.

**Evaluation** Following the analysis of the EL graph, it will be imperative to evaluate the results and examine the usefulness of SugarCube's quantification method. One way to accomplish this task is to study the correlation between the percolation level of various topics, specially global ones, and observable trends in the Blogosphere obtained using tools such as buzz trackers, top-link classifiers, topic detectors and other established and widely accepted methods.

**Prediction** The next major step in this research is to enhance the model so that it can predict the likelihood of a topic in becoming global. We will investigate integrating various approaches such as (Liben-Nowell and Kleinberg 2003) with definition of Percolation Threshold and our approach to keep BL and EL layers separate.

**Non-blog Domain** The Sugar Cube model uses the characteristics of a relationship network (BL) in order to deduce the level of influence and propagation of ideas that are frequently published (blog posts in EL). In this sense, there are other domains that share similar characteristics with the Blogosphere: The network of academic citations consists of a relationship network (co-authorship) and frequently published work (academic papers). Similar qualities might also be found in the network of industrial patents.

## 8. Acknowledgments

Special thanks to Microsoft Research Cambridge for generously funding part of this research, and to Information Retrieval Facility (IRF) for providing the computing infrastructure that made the experiments possible.

## References

Adar, E., and Adamic, L. 2005. Tracking information epidemics in blogspace. *WI '05: Proceedings of the 2005 IEEE/WIC/ACM International Conference on Web Intelligence*.

Ahmed, E., and Abdusalam, H. 2000. On social percolation and small world network. *The European Physical Journal B-Condensed Matter*.

Ante, S. E.; Grover, R.; and Green, H. 2007. In search of MyProfits. *BusinessWeek*.

Bethe, H. A. 1935. Statistical theory of superlattices. In *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, volume 150 (No 871), 552–575. The Royal Society.

Braga, G. A.; Sanchis, R.; and Schieber, T. A. 2005. Critical percolation on a bethe lattice revisited. *SIAM review* 47:349–365.

Efros, A. L. 1986. *Physics and Geometry of Disorder. Percolation Theory*. Mir Publishers Moscow.

Essam, J. 1980. Percolation theory. *Rep. Prog. Phys.*

Frisch, H. L., and Hammersley, J. M. 1963. Percolation processes and related topics. *Journal of the Society for Industrial and Applied Mathematics* 11(4):894–918.

Herring, S.; Scheidt, L.; Bonus, S.; and Wright, E. 2004. Bridging the gap: A genre analysis of weblogs. *System Sciences, 2004. Proceedings of the 37th Annual Hawaii International Conference on* 11.

ICWSM. 2009. ICWSM 2009 Spinn3r Dataset. In *Proceedings of the Third International Conference on Weblogs and Social Media (ICWSM 2009)*.

Jeraldo, P. 2008. Critical phenomena in percolation theory. *guava.physics.uiuc.edu*.

Kale, A. 2007. Modeling trust and influence on blogosphere using link polarity.

Kaplan, D. 2008. Social networks and profitability: Facebook, MySpace look beyond ads. *The Washington Post*.

Kelleher, K. 2008. MySpace and Friends Need to Make Money. And fast. *Wired Magazine*.

Kesten, H. 1982. *Percolation Theory for Mathematics*. Birkhauser.

Leskovec, J., and McGlohon, M. 2007. Cascading behavior in large blog graphs. In *SDM 2007*.

Liben-Nowell, D., and Kleinberg, J. 2003. The link prediction problem for social networks. *CIKM '03: Proceedings of the twelfth international conference on Information and knowledge management*.

Lin, Y.-R.; Sundaram, H.; Chi, Y.; Tatemura, J.; and Tseng, B. 2007. Blog community discovery and evolution based on mutual awareness expansion. *WI '07: Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence*.

Llora, X.; Yasui, N.; Welge, M.; and Goldberg, D. Analyzing trends in the blogosphere using human-centered analysis and visualization tools. *Urbana* 51:61801.

Stauffer, A. A. D. 1994. *Introduction to Percolation Theory*. CRC.

Sykes, M. F., and Essam, J. W. 1963. Some exact critical percolation probabilities for bond and site problems in two dimensions. *Phys. Rev. Lett. 10, 3 - 4*.

Takahashi, K. 2004. An application of percolation theory to political science. *necsi.org*.

## Notes

1. There is not enough information to make a certain judgement on this URL.